

Newton's method on Graßmann manifolds

Uwe Helmke · Knut Hüper · Jochen Trumpf*

Received: date / Accepted: date

Abstract A general class of Newton algorithms on Graßmann and Lagrange–Graßmann manifolds is introduced, that depends on an arbitrary pair of local coordinates. Local

* corresponding author

The first author was partially supported by a grant from BMBF within the FprofUnd programme. The second and third author are with National ICT Australia Limited which is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's Ability* and the ICT Research Centre of Excellence Program.

U. Helmke
Mathematisches Institut
Universität Würzburg
Am Hubland
97074 Würzburg
Germany
E-mail: helmke@mathematik.uni-wuerzburg.de

K. Hüper
National ICT Australia Limited
Canberra Research Laboratory
Locked Bag 8001
Canberra ACT 2601
Australia
E-mail: knut.hueper@nicta.com.au
and
Department of Information Engineering
The Australian National University

J. Trumpf
Department of Information Engineering
The Australian National University
Canberra ACT 0200
Australia
Tel.: +61-2-61258677
Fax: +61-2-61258660
E-mail: jochen.trumpf@anu.edu.au
and
National ICT Australia Limited
Canberra Research Laboratory

quadratic convergence of the algorithm is shown under a suitable condition on the choice of coordinate systems. Our result extends and unifies previous convergence results for Newton's method on a manifold. Using special choices of the coordinates, new numerical algorithms are derived for principal component analysis and invariant subspace computations with improved computational complexity properties.

Keywords Newton's method · Graßmann and Lagrange Graßmann manifold · smooth parametrizations · Riemannian metric

Mathematics Subject Classification (2000) MSC 49M15 · MSC 53B20 · MSC 65F15 · MSC 15A18

1 Introduction

Riemannian optimization is a relatively recent approach towards constrained optimization that uses full information on the underlying geometry of the constraint set in order to set up the optimization algorithms. The method is particularly useful if the basic ingredients from differential geometry, such as the Levi-Civita connection and geodesics are explicitly available. This happens in many application problems arising in signal processing and numerical linear algebra, where optimization naturally takes place on homogeneous spaces, such as e.g. Stiefel or Graßmann manifolds. In this paper, we describe a new class of Newton algorithms on Graßmann manifolds and study applications to eigenvalue and invariant subspace computations.

The idea of using differential geometric methods to construct gradient descent algorithms for constrained optimization on smooth manifolds is of course not new and we refer to the textbooks [11, 6, 16] for further information. Such gradient algorithms use first order derivative information on the function and thus can be described in a rather straightforward way. In contrast, Newton's method on a manifold requires second order information on the function, using an affine connection in order to define the Hessian. This can be done in several different ways, thus leading to a variety of possible implementations of the Newton algorithm.

In D. Gabay's work [5], the intrinsic Newton method on a Riemannian manifold is defined via the Levi-Civita connection, taking iteration steps along associated geodesics. More generally, M. Shub [14] proposed a Newton method to compute a zero of a smooth vector field on a smooth manifold endowed with an affine connection. His algorithm is defined for arbitrary families of smooth projections $\pi_p : T_p M \rightarrow M, p \in M$, from the tangent bundle which have derivative equal to the identity at the base point. Therefore it is more general than Gabay's method and can be employed on arbitrary manifolds, without having to specify a Riemannian metric. In the case of a gradient vector field on a Riemannian manifold endowed with the Levi-Civita connection, Shub's algorithm coincides with Gabay's, when $\{\pi_p\}_{p \in M}$ are the Riemannian normal coordinates.

In the PhD theses of St. Smith and R. Mahony [15, 13], see also [4], the Newton method along geodesics of Gabay [5] was rediscovered. However, the convergence proofs developed in these papers do not apply to the more general situation studied by Shub, except for the special case of Rayleigh quotient optimization on the unit sphere. In his recent PhD Thesis, P.-A. Absil [1], see also [2], further discusses the Newton method along geodesics and derives a cubic convergence result in a special case. Moreover, variants with different projections were proposed, too. There are many more, recent

publications discussing aspects of Newton methods on Riemannian manifolds. We want to specifically mention the paper by Adler et al. [3] which is similar in spirit to this paper in so far as it provides explicit formulas for parametrizations and Newton algorithms on $(\text{SO}_3)^N$.

In this paper, we propose a general approach to Newton's method on both Graßmann and Lagrange Graßmann manifolds that incorporates the previous ones as special cases, but allows also for implementations with improved computational complexity. We do so by replacing the family of smooth projections by an arbitrary pair of local coordinates μ_p, ν_p with equal derivatives $D\mu_p(0) = D\nu_p(0)$. Although this generalization might look minor at first sight, it is actually crucial to achieve better performance. Following [7] and extending the known local quadratic convergence result for the intrinsic Riemannian Newton method, we prove local quadratic convergence of the generalized Newton algorithm. The Newton method on the Lagrange Graßmannian has not been considered before, but has important applications in control (e.g. to algebraic Riccati equations in linear quadratic control).

The paper is structured as follows. In order to enhance the readability of the paper for non-experts, we begin with a brief summary of the basic differential geometry of the classical Graßmann manifold and the Lagrange Graßmannian, respectively, deriving explicit formulas for (projections onto) tangent spaces, normal spaces, gradients, Hessians, and geodesics. We then compute the Riemannian normal coordinates of the two types of Graßmannians. Using approximations of the exponential map via e.g. Padé approximants or the QR factorization, then leads to alternative coordinate systems and resulting simplified implementations of the Newton algorithm. By generalizing the construction of Shub, we introduce the Newton algorithm via a pull back/push forward scheme defined by an arbitrary pair of local coordinates for the Graßmannians. This leads to a rich family of intrinsically defined Newton methods that have potential for considerable computational advantages compared with the previously known algorithms. In fact, instead of relying upon the use of Riemannian normal coordinates, that are difficult to compute with, we advocate to use the much more easily computable local coordinates via the QR -factorization.

For example, in Edelman et al. [4] the steps of the Newton algorithm on the classical Graßmannian are defined in the ambient Euclidean space of the associated Stiefel manifold. This leads them to solving sequences of Sylvester equations in higher dimensional matrix spaces than necessary. In contrast, our algorithms works with the minimal number of parameters, given by the dimension of the Graßmannian. Moreover, our algorithms do not require the iterative calculation of matrix exponentials, but only involve finite step iterations using efficient QR -computations.

Finally, we apply these techniques to eigenspace computations. By applying our Newton scheme to the Rayleigh quotient function on the Graßmann (and Lagrange Graßmann) manifold, we obtain a new class of iterative algorithms for principal component analysis with improved computational complexity. For eigenspace computations of arbitrary, not necessarily symmetric, matrices we derive an apparently new class of Newton algorithms, that requires the repeated computations of solutions to nested Sylvester type equations.

2 Riemannian geometry of the Graßmann manifold

In this section we describe the basics for the Riemannian geometry of Graßmann manifolds, i.e. tangent and normal spaces, Riemannian metrics and geodesics. We focus on the real Graßmannian; the results carry through *mutatis mutandis* for complex Graßmannians, too.

Recall, that the Graßmann manifold $\text{Gr}_{m,n}$ is defined as the set of m -dimensional \mathbb{R} -linear subspaces of \mathbb{R}^n . It is a smooth, compact manifold of dimension $m(n - m)$ and provides a natural generalization of the familiar projective spaces. Let denote

$$\text{O}_n := \{X \in \mathbb{R}^{n \times n} \mid X^\top X = I\}. \quad (2.1)$$

and

$$\text{SO}_n := \{X \in \text{O}_n \mid \det X = 1\} \quad (2.2)$$

The Graßmann manifold can also be viewed in an equivalent way as a homogeneous space $\text{SO}_n(\mathbb{R})/H$, cf. e.g. [6] and see below for a definition of H , for the transitive SO_n -action

$$\begin{aligned} \sigma : \text{SO}_n \times \text{Gr}_{m,n} &\rightarrow \text{Gr}_{m,n}, \\ (T, V) &\mapsto TV. \end{aligned} \quad (2.3)$$

Let

$$\mathbf{V}_0 = \text{colspan} \begin{bmatrix} I_m \\ 0 \end{bmatrix} \in \text{Gr}_{m,n} \quad (2.4)$$

denote the standard m -dimensional subspace of \mathbb{R}^n that is spanned by the first m standard basis vectors of \mathbb{R}^n . Then the stabilizer subgroup $H := \text{Stab}(\mathbf{V}_0)$ of \mathbf{V}_0 is given by

$$H = \left\{ \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \in \text{SO}_n \mid U \in \text{O}_m, V \in \text{O}_{n-m} \right\}, \quad (2.5)$$

i.e. by the compact Lie subgroup of SO_n consisting of all block diagonal orthogonal matrices. The map

$$\text{SO}_n / H \rightarrow \text{Gr}_{m,n}, \quad \Theta H \mapsto \Theta \mathbf{V}_0 \quad (2.6)$$

then defines a diffeomorphism of the Graßmann manifold with the homogeneous space SO_n / H . See Edelman et al. [4], Absil [1] and Hüper and Trumpf [7] for further details on Newton's method on $\text{Gr}_{m,n}$, in a variant that exploits the homogeneous space structure of the Graßmann manifold. Here we develop a different approach, by identifying $\text{Gr}_{m,n}$ with a set of self-adjoint projection operators.

Thus we define the *Graßmannian* as

$$\text{Gr}_{m,n} := \{P \in \mathbb{R}^{n \times n} \mid P^\top = P, P^2 = P, \text{tr } P = m\}, \quad (2.7)$$

the manifold of rank m symmetric projection operators of \mathbb{R}^n ; see [6] for the construction of a natural bijection with the Graßmann manifold and a proof that it defines a diffeomorphism. In the sequel we will describe the Riemannian geometry directly for the submanifold $\text{Gr}_{m,n}$ of $\mathbb{R}^{n \times n}$. As we will see, this approach has advantages that simplify both the analysis and design of Newton-based algorithms for the computation of principal components.

We begin by recalling the following known and basic fact on the Graßmannian; see [6, Section 2.1] for a proof in the more general context of isospectral manifolds. Let

$$\text{Sym}_n := \{S \in \mathbb{R}^{n \times n} \mid S^\top = S\} \quad (2.8)$$

and

$$\mathfrak{so}_n := \{\Omega \in \mathbb{R}^{n \times n} \mid \Omega^\top = -\Omega\} \quad (2.9)$$

denote the vector spaces of real symmetric and real skew-symmetric matrices, respectively.

Theorem 2.1 (a) *The Graßmannian $\text{Gr}_{m,n}$ is a smooth, compact submanifold of Sym_n of dimension $m(n-m)$.*
 (b) *The tangent space of $\text{Gr}_{m,n}$ at an element $P \in \text{Gr}_{m,n}$ is given as*

$$T_P \text{Gr}_{m,n} = \{[P, \Omega] \mid \Omega \in \mathfrak{so}_n\}. \quad (2.10)$$

Here $[P, \Omega] := P\Omega - \Omega P$ denotes the matrix commutator (Lie bracket).

Let

$$\begin{aligned} \text{ad}_P : \mathbb{R}^{n \times n} &\rightarrow \mathbb{R}^{n \times n}, \\ \text{ad}_P(X) &:= [P, X] \end{aligned} \quad (2.11)$$

denote the adjoint representation at P . For a projection operator P it enjoys the following property.

Lemma 2.1 *For any $P \in \text{Gr}_{m,n}$, the minimal polynomial of $\text{ad}_P : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ is equal to $s^3 - s$. Thus $\text{ad}_P^3 = \text{ad}_P$. Moreover,*

$$\text{ad}_P^2 X = [P, [P, X]] = X \quad (2.12)$$

holds for all tangent vectors $X \in T_P \text{Gr}_{m,n}$.

Proof From $P^2 = P$ we get

$$\text{ad}_P^2 X = [P, [P, X]] = P^2 X + X P^2 - 2PXP = PX + XP - 2PXP \quad (2.13)$$

and therefore, using $P^2 = P$ again

$$\begin{aligned} \text{ad}_P^3 X &= P(PX + XP - 2PXP) - (PX + XP - 2PXP)P \\ &= PX - XP \\ &= \text{ad}_P X \end{aligned} \quad (2.14)$$

for all $n \times n$ -matrices X . If $X = [P, \Omega]$ is a tangent vector, then $\text{ad}_P^2 X = \text{ad}_P^3 \Omega = \text{ad}_P \Omega = X$. The result follows. \square

We use this result to describe the normal bundle of $\text{Gr}_{m,n}$. In the sequel, we will always endow Sym_n with the Frobenius inner product, defined by

$$\langle X, Y \rangle := \text{tr}(XY) \quad (2.15)$$

for all $X, Y \in \text{Sym}_n$. Since the tangent space $T_P \text{Gr}_{m,n} \subset \text{Sym}_n$ is a subset of Sym_n (using the usual identification of $T_P \text{Sym}_n$ with Sym_n), we can define the normal space at P to be the vector space

$$N_P \text{Gr}_{m,n} = (T_P \text{Gr}_{m,n})^\perp := \{X \in \text{Sym}_n \mid \text{tr}(XY) = 0 \text{ for all } Y \in T_P \text{Gr}_{m,n}\}. \quad (2.16)$$

Proposition 2.1 *Let $P \in \text{Gr}_{m,n}$ be arbitrary.*

1. *The normal subspace in Sym_n is given as*

$$N_P \text{Gr}_{m,n} = \{X - \text{ad}_P^2 X \mid X \in \text{Sym}_n\}. \quad (2.17)$$

2. *The linear map*

$$\pi : \text{Sym}_n \rightarrow \text{Sym}_n, \quad X \mapsto \text{ad}_P^2 X = [P, [P, X]] \quad (2.18)$$

is the self-adjoint projection operator onto $T_P \text{Gr}_{m,n}$ with kernel $N_P \text{Gr}_{m,n}$.

Proof For any tangent vector $[P, \Omega] \in T_P \text{Gr}_{m,n}$, where $\Omega^\top = -\Omega$, and any $X = X^\top$, we have

$$\begin{aligned} \text{tr}([P, \Omega](X - \text{ad}_P^2 X)) &= \text{tr}([X, P] - [\text{ad}_P^2 X, P])\Omega \\ &= \text{tr}([X, P] + \text{ad}_P^3 X)\Omega \\ &= \text{tr}((\text{ad}_P^3 X - \text{ad}_P X)\Omega) \\ &= 0, \end{aligned} \quad (2.19)$$

since $\text{ad}_P^3 = \text{ad}_P$. Therefore, $T_P \text{Gr}_{m,n}$ and $\{X - \text{ad}_P^2 X \mid X \in \text{Sym}_n\}$ are orthogonal subspaces of Sym_n with respect to the Frobenius inner product. Their sum also spans Sym_n , as otherwise there exists a nontrivial $S \in \text{Sym}_n$ that is orthogonal to both spaces; but then for all $\Omega \in \mathfrak{so}_n$

$$\text{tr}(S[P, \Omega]) = \text{tr}([S, P]\Omega) = 0 \implies [S, P] = 0, \quad (2.20)$$

and for all $X \in \text{Sym}_n$, using (2.20)

$$\text{tr}(S(X - \text{ad}_P^2 X)) = \text{tr}(SX - [S, P][P, X]) = \text{tr}(SX) = 0 \quad (2.21)$$

which implies $S = 0$, a contradiction. Thus the two spaces define an orthogonal sum decomposition of Sym_n and therefore $\{X - \text{ad}_P^2 X \mid X \in \text{Sym}_n\}$ must be the normal space. This completes the proof for the first claim.

Since $\pi = \text{ad}_P^2$, we have

$$\pi^2 = \text{ad}_P^4 = \text{ad}_P^2 = \pi \quad (2.22)$$

because $\text{ad}_P^3 = \text{ad}_P$. Moreover, by definition of π we have $\text{im } \pi \subset T_P \text{Gr}_{m,n}$, cf. (2.10), and for any $X \in T_P \text{Gr}_{m,n}$ we have by Lemma 2.1 that $\pi(X) = X$. Therefore

$$\text{im } \pi = T_P \text{Gr}_{m,n}. \quad (2.23)$$

For any $X - \text{ad}_P^2 X \in N_P \text{Gr}_{m,n}$ we have

$$\pi(X - \text{ad}_P^2 X) = \text{ad}_P^2 X - \text{ad}_P^4 X = 0, \quad (2.24)$$

by (2.22). Since $N_P \text{Gr}_{m,n}$ is the orthogonal complement to the tangent space in Sym_n , a straight forward dimension argument yields $\ker \pi = N_P \text{Gr}_{m,n}$. Finally, using the Frobenius inner product on Sym_n , we have for all $X_1, X_2 \in \text{Sym}_n$

$$\begin{aligned} \langle \pi(X_1), X_2 \rangle &= \text{tr}((\text{ad}_P^2 X_1)X_2) \\ &= \text{tr}([P, [P, X_1]]X_2) \\ &= \text{tr}([P, [P, X_2]]X_1) \\ &= \langle X_1, \pi(X_2) \rangle. \end{aligned} \quad (2.25)$$

Thus π is self-adjoint and the result follows. \square

A formula for π in the language of linear maps has already been given in [12, Section 4.2].

There are at least two natural Riemannian metrics defined on the Grassmannian $\text{Gr}_{m,n}$, the induced *Euclidean metric* and the *normal metric*, cf. e.g. [6] or [13].

The Euclidean Riemannian metric on $\text{Gr}_{m,n}$ is defined by the Frobenius inner product on the tangent spaces

$$\langle X, Y \rangle := \text{tr}(XY) \quad (2.26)$$

for all $X, Y \in T_P \text{Gr}_{m,n}$ which is induced by the embedding space Sym_n .

The normal Riemannian metric has a somewhat more complicated definition. Consider the surjective linear map

$$\begin{aligned} \text{ad}_P : \mathfrak{so}_n &\rightarrow T_P \text{Gr}_{m,n}, \\ \Omega &\mapsto [P, \Omega] \end{aligned} \quad (2.27)$$

with kernel

$$\ker \text{ad}_P = \{\Omega \in \mathfrak{so}_n \mid P\Omega = \Omega P\}. \quad (2.28)$$

We regard \mathfrak{so}_n as an inner product space, endowed with the Frobenius inner product $\langle \Omega_1, \Omega_2 \rangle = \text{tr}(\Omega_1^\top \Omega_2) = -\text{tr}(\Omega_1 \Omega_2)$. Then ad_P induces an isomorphism of vector spaces

$$\widehat{\text{ad}}_P : (\ker \text{ad}_P)^\perp \rightarrow T_P \text{Gr}_{m,n} \quad (2.29)$$

and therefore induces an isometry of inner product spaces, by defining an inner product on $T_P \text{Gr}_{m,n}$ via

$$\langle \langle X, Y \rangle \rangle_P := -\text{tr}(\widehat{\text{ad}}_P^{-1}(X) \widehat{\text{ad}}_P^{-1}(Y)). \quad (2.30)$$

Note, that this inner product on $T_P \text{Gr}_{m,n}$, called the normal Riemannian metric, might vary with the basepoint P . Luckily, the situation is better than one would expect, as Proposition 2.3 below shows.

But first we will show that the operator ad_P^2 , $P \in \text{Gr}_{m,n}$, is equally well behaved on \mathfrak{so}_n as it is on Sym_n , cf. Proposition 2.1.

Proposition 2.2 *Let $P \in \text{Gr}_{m,n}$ be arbitrary. The linear map*

$$\text{ad}_P^2 : \mathfrak{so}_n \rightarrow \mathfrak{so}_n, \quad \Omega \mapsto [P, [P, \Omega]] \quad (2.31)$$

is the self-adjoint projection operator onto $(\ker \text{ad}_P)^\perp$ along $\ker \text{ad}_P$.

Proof Let $\Omega \in \mathfrak{so}_n$ be arbitrary. By Lemma 2.1 we know that $X := \text{ad}_P(\Omega) = \text{ad}_P(\text{ad}_P^2 \Omega)$. But since for all $\Omega_1 \in \ker \text{ad}_P \subset \mathfrak{so}_n$

$$\text{tr}(\text{ad}_P^2(\Omega) \Omega_1^\top) = \text{tr}(\text{ad}_P(\Omega) \text{ad}_P(\Omega_1)) = 0, \quad (2.32)$$

we conclude $\text{ad}_P^2 \Omega \in (\ker \text{ad}_P)^\perp \subset \mathfrak{so}_n$. Now let $\Omega \in (\ker \text{ad}_P)^\perp$. Then $\text{ad}_P^2 \Omega \in (\ker \text{ad}_P)^\perp$ and hence $\Omega - \text{ad}_P^2 \Omega \in (\ker \text{ad}_P)^\perp$. By Lemma 2.1 $\text{ad}_P(\Omega - \text{ad}_P^2 \Omega) = \text{ad}_P \Omega - \text{ad}_P^3 \Omega = 0$ and hence $\Omega - \text{ad}_P^2 \Omega \in \ker \text{ad}_P$. It follows $\Omega - \text{ad}_P^2 \Omega = 0$ and thus $\text{ad}_P^2 \Omega = \Omega$.

We have shown $\text{im } \text{ad}_P^2 \subset (\ker \text{ad}_P)^\perp$ and that the restriction of ad_P^2 to $(\ker \text{ad}_P)^\perp$ is the identity. It remains to show that $\ker \text{ad}_P^2 = \ker \text{ad}_P$, but this follows readily from Lemma 2.1. \square

Proposition 2.3 *The Euclidean and normal Riemannian metrics on the Graßmannian $\text{Gr}_{m,n}$ coincide, i.e. for all $P \in \text{Gr}_{m,n}$ and for all $X, Y \in T_P \text{Gr}_{m,n}$ we have*

$$\text{tr}(X^\top Y) = -\text{tr}\left(\widehat{\text{ad}}_P^{-1}(X) \widehat{\text{ad}}_P^{-1}(Y)\right). \quad (2.33)$$

Proof Choose

$$\Omega_1, \Omega_2 \in (\ker \text{ad}_P)^\perp \text{ with } X = [P, \Omega_1] \text{ and } Y = [P, \Omega_2]. \quad (2.34)$$

Then

$$-\text{tr}(\widehat{\text{ad}}_P^{-1}(X) \widehat{\text{ad}}_P^{-1}(Y)) = \text{tr}(\Omega_1^\top \Omega_2). \quad (2.35)$$

On the other hand

$$\begin{aligned} \text{tr}(X^\top Y) &= \text{tr}([P, \Omega_1][P, \Omega_2]) \\ &= \text{tr}([P, [P, \Omega_1]]^\top \Omega_2). \end{aligned} \quad (2.36)$$

Now by Proposition 2.2 we know that $\text{ad}_P^2 \Omega_1 = \Omega_1$ and this implies

$$\text{tr}(X^\top Y) = \text{tr}(\Omega_1^\top \Omega_2), \quad (2.37)$$

as claimed. \square

Since these two Riemannian metrics on the Graßmannian coincide, they also define the same geodesics. Thus, in the sequel, we focus on the Euclidean metric. Note, that the above result is not true for arbitrary flag manifolds and in fact, the geodesics are then different for the two metrics. The following result characterizes the geodesics on $\text{Gr}_{m,n}$.

Theorem 2.2 *The geodesics of $\text{Gr}_{m,n}$ are exactly the solutions of the second order differential equation*

$$\ddot{P} + [\dot{P}, [\dot{P}, P]] = 0. \quad (2.38)$$

The unique geodesic $P(t)$ with initial conditions $P(0) = P_0 \in \text{Gr}_{m,n}$, $\dot{P}(0) = \dot{P}_0 \in T_{P_0} \text{Gr}_{m,n}$ is given by

$$P(t) = e^{t[\dot{P}_0, P_0]} P_0 e^{-t[\dot{P}_0, P_0]}. \quad (2.39)$$

Proof The geodesics of $\text{Gr}_{m,n}$ for the Euclidean metric are characterized as the curves $P(t) \in \text{Gr}_{m,n}$, such that $\dot{P}(t)$ is a normal vector for all $t \in \mathbb{R}$. This condition is equivalent to the existence of $S(t) = S(t)^\top$ with

$$\ddot{P} = S - \text{ad}_P^2 S. \quad (2.40)$$

Since by Lemma 2.1 $\text{ad}_P^3 = \text{ad}_P$ this implies

$$\text{ad}_P \ddot{P} = [P, \ddot{P}] = 0. \quad (2.41)$$

Moreover, any curve $P(t) \in \text{Gr}_{m,n}$ satisfies the identity

$$\dot{P} = \text{ad}_P^2 \dot{P}, \quad (2.42)$$

as ad_P^2 acts as the identity on the tangent space $T_P \text{Gr}_{m,n}$. By differentiating equation (2.42) we obtain

$$\begin{aligned}\ddot{P} &= [P, [P, \ddot{P}]] + [P, [\dot{P}, \dot{P}]] + [\dot{P}, [P, \dot{P}]] \\ &= -\text{ad}_P^2 P + \text{ad}_P^2 \ddot{P}.\end{aligned}\tag{2.43}$$

Therefore, if $P(t)$ is a geodesic, then $\text{ad}_P^2(\ddot{P}) = 0$ and

$$\begin{aligned}\ddot{P} &= -\text{ad}_P^2 P + \text{ad}_P^2 \ddot{P} \\ &= -\text{ad}_P^2 P\end{aligned}\tag{2.44}$$

and therefore satisfies $\ddot{P} + [\dot{P}, [\dot{P}, P]] = 0$, as claimed. We now check, that every curve $P(t)$ as in (2.39) is a solution to (2.38). Let $\Omega := [\dot{P}_0, P_0]$. Then

$$\dot{P} = [\Omega, P], \quad \ddot{P} = [\Omega, [\Omega, P]]\tag{2.45}$$

and thus (2.38) is equivalent to

$$[\Omega, [\Omega, P]] + [[\Omega, P], [[\Omega, P], P]] = 0.\tag{2.46}$$

Multiplying by the left and right with $e^{-t\Omega}$ and $e^{t\Omega}$ respectively, we see that (2.46) is equivalent to

$$[\Omega, [\Omega, P_0]] + [[\Omega, P_0], [[\Omega, P_0], P_0]] = 0.\tag{2.47}$$

Without loss of generality we can assume that

$$P_0 = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}\tag{2.48}$$

and therefore

$$\Omega = \begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix}\tag{2.49}$$

with $Z \in \mathbb{R}^{m \times (n-m)}$. Thus

$$[\Omega, [\Omega, P_0]] = \begin{bmatrix} -2ZZ^\top & 0 \\ 0 & 2Z^\top Z \end{bmatrix}\tag{2.50}$$

and also

$$[[\Omega, P_0], [[\Omega, P_0], P_0]] = \begin{bmatrix} 2ZZ^\top & 0 \\ 0 & -2Z^\top Z \end{bmatrix}.\tag{2.51}$$

This implies (2.46) and shows that any curve given by (2.39) is a solution of (2.38). Since any $P_0 \in \text{Gr}_{m,n}$ and $[\dot{P}_0, P_0] \in T_P \text{Gr}_{m,n}$ are admissible initial conditions for (2.38), and since the resulting initial value problem has a unique solution (namely (2.39)), this shows that (2.39) is exactly the set of *all* solutions of (2.38). Moreover, for the particular initial point

$$P_0 = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}\tag{2.52}$$

one observes that

$$[[\Omega, P_0], [[\Omega, P_0], P_0]] = \begin{bmatrix} 2ZZ^\top & 0 \\ 0 & -2Z^\top Z \end{bmatrix}\tag{2.53}$$

is a normal vector to the Graßmannian at P_0 . Thus, by invariance of the normal bundle under orthogonal similarity transformations $P_0 \mapsto \Theta^\top P_0 \Theta$, $\Theta \in \text{SO}_n$, we see that $[\dot{P}, [\dot{P}, P]]$ is a normal vector to $T_P \text{Gr}_{m,n}$ for all $\dot{P} \in T_P \text{Gr}_{m,n}$. Thus, for any solution $P(t)$ of (2.38) also $\dot{P} = -[\dot{P}, [\dot{P}, P]]$ is a normal vector, and hence all solutions of (2.38) are geodesics. \square

The above explicit formula for geodesics leads to the following formula for the geodesic distance between two points on a Graßmannian. We omit the simple proof; see also [1] for a slightly different formula which is only valid on an open and dense subset of the Graßmannian.

Corollary 2.1 *Let $P, Q \in \text{Gr}_{m,n}$. Given any $\Theta \in \text{SO}_n$ such that*

$$P = \Theta^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta$$

we define

$$\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & Q_{22} \end{bmatrix} := \Theta Q \Theta^\top.$$

Let $1 \geq \lambda_1 \geq \dots \geq \lambda_m \geq 0$ denote the eigenvalues of Q_{11} . The geodesic distance of P to Q in $\text{Gr}_{m,n}$ is given by

$$\text{dist}(P, Q) = \sqrt{2 \sum_{i=1}^m \arccos^2(\sqrt{\lambda_i})}. \quad (2.54)$$

Alternatively, let $1 \geq \mu_1 \geq \dots \geq \mu_{n-m} \geq 0$ denote the eigenvalues of Q_{22} . Then

$$\text{dist}(P, Q) = \sqrt{2 \sum_{i=1}^{n-m} \arcsin^2(\sqrt{\mu_i})}. \quad (2.55)$$

In particular, if $P, Q \in \text{Gr}_{m,n}$ with $Q = YY^\top$, $Y^\top Y = I_m$, then

$$\frac{1}{2} \text{dist}^2(P, Q) = \text{tr} \left(\arccos^2((Y^\top P Y)^{\frac{1}{2}}) \right). \quad (2.56)$$

Note that formula (2.55) is more efficient in the case $2m > n$. Note also that our formulas imply that the maximal length of a simple closed geodesic in $\text{Gr}_{m,n}$ is $\sqrt{2m} \cdot \pi$ for $2m \leq n$ and $\sqrt{2(n-m)} \cdot \pi$ for $2m > n$.

2.1 Parametrizations and Coordinates for the Graßmannian

In this section we briefly recall the notion of local parametrization for smooth manifolds. For further details we refer to [10]. Let M be a smooth n -dimensional real manifold then for every point $p \in M$ there exists a smooth map

$$\mu_p : \mathbb{R}^n \longrightarrow M, \quad \mu_p(0) = p$$

which is a local diffeomorphism around $0 \in \mathbb{R}^n$. Such a map is called a *local parametrization around p* .

We consider local parametrizations for the Graßmannian via the tangent space, i.e. families of smooth maps

$$\mu_P : T_P \text{Gr}_{m,n} \rightarrow \text{Gr}_{m,n} \quad (2.57)$$

satisfying

$$\mu_P(0) = P \text{ and } D\mu_P(0) = \text{id}. \quad (2.58)$$

We introduce three different choices of such local parametrizations.

2.1.1 Riemannian normal coordinates

Riemannian normal coordinates are defined through the Riemannian exponential map (see e.g. [8])

$$\begin{aligned} \mu_P^{\text{exp}} &= \exp_P : T_P \text{Gr}_{m,n} \rightarrow \text{Gr}_{m,n}, \\ \exp_P(\xi) &= e^{[\xi, P]} P e^{-[\xi, P]}. \end{aligned} \quad (2.59)$$

Remark 2.1 Note that by Theorem 2.2 the unique geodesic $P(t)$ with initial conditions $P(0) = P_0$ and $\dot{P}(0) = \dot{P}_0$ is given by $P(t) = \exp_{P_0}(t\dot{P}_0) = \mu_{P_0}^{\text{exp}}(t\dot{P}_0)$.

Obviously, \exp_P is smooth with

$$\exp_P(0) = P \quad (2.60)$$

and

$$D\exp_P(0) = \text{id}, \quad (2.61)$$

as

$$\begin{aligned} D\exp_P(0)\xi &= [[\xi, P], P] \\ &= \text{ad}_P^2 \xi \\ &= \xi \quad \text{for all } \xi \in T_P \text{Gr}_{m,n}. \end{aligned} \quad (2.62)$$

Such Riemannian normal coordinates can be explicitly computed as follows. Given any $\Theta \in \text{SO}_n$ with

$$P = \Theta^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta \quad (2.63)$$

we can write

$$[\xi, P] = \Theta^\top \begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix} \Theta \quad (2.64)$$

with $Z \in \mathbb{R}^{m \times (n-m)}$. Since

$$[\xi, P]^{2m} = \Theta^\top \begin{bmatrix} (-ZZ^\top)^m & 0 \\ 0 & (-Z^\top Z)^m \end{bmatrix} \Theta \quad (2.65)$$

we obtain

$$e^{[\xi, P]} = \Theta^\top e^{\begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix}} \Theta = \Theta^\top \begin{bmatrix} \cos \sqrt{ZZ^\top} & Z \frac{\sin \sqrt{Z^\top Z}}{\sqrt{Z^\top Z}} \\ -\frac{\sin \sqrt{Z^\top Z}}{\sqrt{Z^\top Z}} Z^\top & \cos \sqrt{Z^\top Z} \end{bmatrix} \Theta. \quad (2.66)$$

Here, as usual, it is understood that

$$\begin{aligned}
Z \frac{\sin \sqrt{Z^\top Z}}{\sqrt{Z^\top Z}} &= Z(Z^\top Z)^{-\frac{1}{2}} \sin(Z^\top Z)^{\frac{1}{2}} \\
&= Z \sum_{i=0}^{\infty} \frac{(-1)^i ((Z^\top Z)^{\frac{1}{2}})^{2i}}{(2i+1)!} \\
&= Z \sum_{i=0}^{\infty} \frac{(-1)^i (Z^\top Z)^i}{(2i+1)!} \\
&= \sum_{i=0}^{\infty} \frac{(-1)^i (ZZ^\top)^i}{(2i+1)!} Z \\
&= \frac{\sin \sqrt{ZZ^\top}}{\sqrt{ZZ^\top}} Z.
\end{aligned} \tag{2.67}$$

Therefore

$$\begin{aligned}
\exp_P(\xi) &= \Theta^\top \begin{bmatrix} \cos \sqrt{ZZ^\top} & \\ -\frac{\sin \sqrt{ZZ^\top}}{\sqrt{ZZ^\top}} Z^\top & \end{bmatrix} \begin{bmatrix} \cos \sqrt{ZZ^\top} - Z \frac{\sin \sqrt{ZZ^\top}}{\sqrt{ZZ^\top}} \\ \end{bmatrix} \Theta \\
&= \Theta^\top \begin{bmatrix} \cos^2 \sqrt{ZZ^\top} & -\cos \sqrt{ZZ^\top} \frac{\sin \sqrt{ZZ^\top}}{\sqrt{ZZ^\top}} Z \\ -Z^\top \frac{\sin \sqrt{ZZ^\top}}{\sqrt{ZZ^\top}} \cos \sqrt{ZZ^\top} & \sin^2 \sqrt{ZZ^\top} \end{bmatrix} \Theta \\
&= \Theta^\top \begin{bmatrix} \cos^2 \sqrt{ZZ^\top} & -\operatorname{sinc}(2\sqrt{ZZ^\top}) Z \\ -Z^\top \operatorname{sinc}(2\sqrt{ZZ^\top}) & \sin^2 \sqrt{ZZ^\top} \end{bmatrix} \Theta \\
&= \frac{1}{2} I_n + \Theta^\top \begin{bmatrix} \frac{1}{2} \cos(2\sqrt{ZZ^\top}) & -\operatorname{sinc}(2\sqrt{ZZ^\top}) Z \\ -Z^\top \operatorname{sinc}(2\sqrt{ZZ^\top}) & -\frac{1}{2} \sin(2\sqrt{ZZ^\top}) \end{bmatrix} \Theta
\end{aligned} \tag{2.68}$$

2.1.2 QR-coordinates

We define QR-coordinates by the map

$$\begin{aligned}
\mu_P^{\text{QR}} : T_P \operatorname{Gr}_{m,n} &\rightarrow \operatorname{Gr}_{m,n}, \\
\xi &\mapsto (I + [\xi, P])_{\mathbb{Q}} P \left((I + [\xi, P])_{\mathbb{Q}} \right)^\top.
\end{aligned} \tag{2.69}$$

Here $M_{\mathbb{Q}}$ denotes the Q -factor in the QR -factorization $M = M_{\mathbb{Q}} M_{\mathbb{R}}$ of M . Note that the matrix

$$I + [\xi, P] = \Theta^\top \begin{bmatrix} I & Z \\ -Z^\top & I \end{bmatrix} \Theta \tag{2.70}$$

is always invertible and therefore the Q -factor $(I + [\xi, P])_{\mathbb{Q}} \in O_n(\mathbb{R})$ exists, and moreover, is unique if the diagonal entries of the upper triangular factor R are chosen positive. From now on, we always choose the R -factor in this way. Actually, the determinant of the Q -factor,

$$\det(I + [\xi, P])_{\mathbb{Q}} = 1, \tag{2.71}$$

i.e., $(I + [\xi, P])_{\mathbb{Q}} \in SO_n(\mathbb{R})$, as it is easily checked that $\det \begin{bmatrix} I & Z \\ -Z^\top & I \end{bmatrix} > 0$ always.

Moreover, by the smoothness of the QR -factorization for general invertible matrices (follows from the Gram-Schmidt procedure rather than from the usual algorithm via Householder transformations), the map μ_P^{QR} is smooth on the tangent spaces $T_P \text{Gr}_{m,n}$ with

$$\mu_P^{\text{QR}}(0) = P. \quad (2.72)$$

Now by (2.72)

$$D\mu_P^{\text{QR}}(0) : T_P \text{Gr}_{m,n} \rightarrow T_P \text{Gr}_{m,n} \quad (2.73)$$

and a straightforward computation shows that $D\mu_P^{\text{QR}}(0) = \text{id}$. In fact, by differentiating the QR -factorization

$$I + t[\xi, P] = Q(t)R(t), \quad Q(0) = I, \quad R(0) = I \quad (2.74)$$

we obtain

$$[\xi, P] = \dot{Q}R + Q\dot{R} \quad (2.75)$$

and therefore at $t = 0$

$$[\xi, P] = \dot{Q}(0) + \dot{R}(0). \quad (2.76)$$

But $[\xi, P]$ and $\dot{Q}(0)$ are skew-symmetric, while $\dot{R}(0)$ is upper triangular. Thus $\dot{R}(0) = 0$ and therefore

$$\left. \frac{d}{dt}(I + t[\xi, P])_Q \right|_{t=0} = \dot{Q}(0) = [\xi, P]. \quad (2.77)$$

This shows

$$D\mu_P^{\text{QR}}(0)\xi = [[\xi, P], P] = \xi \quad (2.78)$$

for all $\xi \in T_P \text{Gr}_{m,n}$, as claimed.

There exist explicit formulas for the Q and R -factors in terms of Cholesky factors. In fact, with

$$P = \Theta^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta \quad (2.79)$$

and since

$$\Theta(I + [\xi, P])\Theta^\top = \begin{bmatrix} I_m & Z \\ -Z^\top & I_{n-m} \end{bmatrix} =: X \quad (2.80)$$

satisfies

$$XX^\top = X^\top X = \begin{bmatrix} I_m + ZZ^\top & 0 \\ 0 & I_{n-m} + Z^\top Z \end{bmatrix} \quad (2.81)$$

the QR -factorization of $X = \begin{bmatrix} I_m & Z \\ -Z^\top & I_{n-m} \end{bmatrix} = X_Q X_R$ is obtained as

$$X_Q = \begin{bmatrix} R_{11}^{-1} & ZR_{22}^{-1} \\ -Z^\top R_{11}^{-1} & R_{22}^{-1} \end{bmatrix} \quad (2.82)$$

and

$$X_R = \begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix}. \quad (2.83)$$

Here R_{11} and R_{22} are the unique Cholesky factors defined by

$$\begin{aligned} R_{11}^\top R_{11} &= I_m + ZZ^\top, \\ R_{22}^\top R_{22} &= I_{n-m} + Z^\top Z. \end{aligned} \quad (2.84)$$

Note, that vanishing of the 12-block of X_R follows from the invertibility of R_{11} and R_{22} and equation (2.81).

2.1.3 Cayley coordinates

Another possibility to introduce easily computable coordinates utilizes the Cayley transform. For any skew-symmetric matrix Ω the *Cayley transform*

$$\begin{aligned} \text{Cay} : \mathfrak{so}_n &\rightarrow \text{SO}_n, \\ \Omega &\rightarrow (2I + \Omega)(2I - \Omega)^{-1} \end{aligned} \quad (2.85)$$

is smooth and satisfies $\text{D Cay}(0) = \text{id}$. The Cayley coordinates are defined as

$$\begin{aligned} \mu_P^{\text{Cay}} : T_P \text{Gr}_{m,n} &\rightarrow \text{Gr}_{m,n}, \\ \xi &\mapsto \text{Cay}([\xi, P]) P \text{Cay}(-[\xi, P]). \end{aligned} \quad (2.86)$$

The above mentioned property of the Cayley transform implies that μ_P^{Cay} is smooth and satisfies

$$\begin{aligned} \mu_P^{\text{Cay}}(0) &= P, \\ \text{D} \mu_P^{\text{Cay}}(0)\xi &= \text{D Cay}(0)([\xi, P])P - P \text{D Cay}(0)[\xi, P] = [[\xi, P], P] = \xi \end{aligned} \quad (2.87)$$

for all tangent vectors $\xi \in T_P \text{Gr}_{m,n}$. Moreover, $\mu_P^{\text{Cay}}(\xi)$ is easily computed as follows. For

$$P = \Theta^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta, \quad \xi = \Theta^\top \begin{bmatrix} 0 & -Z \\ -Z^\top & 0 \end{bmatrix} \Theta \quad (2.88)$$

a straightforward computation shows, using Schur complements and properties of the von Neumann series, that

$$\begin{aligned} \text{Cay} \left(\begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix} \right) &= \begin{bmatrix} 2I_m & Z \\ -Z^\top & 2I_{n-m} \end{bmatrix} \begin{bmatrix} 2I_m & -Z \\ Z^\top & 2I_{n-m} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 2I_m & Z \\ -Z^\top & 2I_{n-m} \end{bmatrix} \begin{bmatrix} \frac{1}{2}(I_m + \frac{1}{4}ZZ^\top)^{-1} & \frac{1}{4}Z(I_{n-m} + \frac{1}{4}Z^\top Z)^{-1} \\ -\frac{1}{4}Z^\top(I_m + \frac{1}{4}ZZ^\top)^{-1} & \frac{1}{2}(I_{n-m} + \frac{1}{4}Z^\top Z)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} I_m - \frac{1}{4}ZZ^\top & Z \\ -Z^\top & I_{n-m} - \frac{1}{4}Z^\top Z \end{bmatrix} \begin{bmatrix} I_m + \frac{1}{4}ZZ^\top & 0 \\ 0 & I_{n-m} + \frac{1}{4}Z^\top Z \end{bmatrix}^{-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mu_P^{\text{Cay}}(\xi) &= \Theta^\top \text{Cay} \left(\begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix} \right) \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \text{Cay} \left(\begin{bmatrix} 0 & -Z \\ Z^\top & 0 \end{bmatrix} \right) \Theta \\ &= \Theta^\top \begin{bmatrix} I_m - \frac{1}{4}ZZ^\top \\ -Z^\top \end{bmatrix} \left(I_m + \frac{1}{4}ZZ^\top \right)^{-2} \begin{bmatrix} I_m - \frac{1}{4}ZZ^\top & -Z \end{bmatrix} \Theta. \end{aligned} \quad (2.89)$$

Now

$$\begin{bmatrix} I_m - \frac{1}{4}ZZ^\top \\ -Z^\top \end{bmatrix} \left(I_m + \frac{1}{4}ZZ^\top \right)^{-1} \quad (2.90)$$

is a basis matrix with orthonormal columns and therefore $\mu_P^{\text{Cay}}(\xi)$ is exactly the projection operator associated with the linear subspace

$$\Theta^\top \text{colspan} \begin{bmatrix} I_m - \frac{1}{4}ZZ^\top \\ -Z^\top \end{bmatrix}. \quad (2.91)$$

2.1.4 Approximation properties of parametrizations

We have already shown that

$$\mathbf{D} \mu_P^{\text{exp}}(0) = \mathbf{D} \mu_P^{\text{Cay}}(0) = \mathbf{D} \mu_P^{\text{QR}}(0) = \text{id}. \quad (2.92)$$

Moreover, it holds

Theorem 2.3 *Let $P \in \text{Gr}_{m,n}$ and $[\xi, P]$ as in (2.63) and (2.64), respectively. Then*

$$\left. \frac{d^2}{d\varepsilon^2} \mu_P^{\text{exp}}(\varepsilon\xi) \right|_{\varepsilon=0} = \left. \frac{d^2}{d\varepsilon^2} \mu_P^{\text{Cay}}(\varepsilon\xi) \right|_{\varepsilon=0} = \left. \frac{d^2}{d\varepsilon^2} \mu_P^{\text{QR}}(\varepsilon\xi) \right|_{\varepsilon=0} = \Theta^\top \begin{bmatrix} -2ZZ^\top & 0 \\ 0 & 2Z^\top Z \end{bmatrix} \Theta. \quad (2.93)$$

Note, that the right hand side of (2.93) is independent of the choice of Θ in (2.63).

Proof Taking derivatives yields

$$\begin{aligned} \left. \frac{d^2}{d\varepsilon^2} \mu_P^{\text{exp}}(\varepsilon\xi) \right|_{\varepsilon=0} &= \left. \frac{d^2}{d\varepsilon^2} e^{[\varepsilon\xi, P]} P e^{-[\varepsilon\xi, P]} \right|_{\varepsilon=0} \\ &= [\xi, P]^2 P + P[\xi, P]^2 - 2[\xi, P]P[\xi, P] \\ &= \Theta^\top \begin{bmatrix} -2ZZ^\top & 0 \\ 0 & 2Z^\top Z \end{bmatrix} \Theta \end{aligned} \quad (2.94)$$

From the theory of Padé approximations it is well known that for all matrices $X \in \mathfrak{so}_n$ and $\varepsilon \in \mathbb{R}$

$$e^{\varepsilon X} = (2I_n + \varepsilon X)(2I_n - \varepsilon X)^{-1} + \mathcal{O}(\varepsilon^3). \quad (2.95)$$

Consequently,

$$\left. \frac{d^2}{d\varepsilon^2} \mu_P^{\text{Cay}}(\varepsilon\xi) \right|_{\varepsilon=0} = \Theta^\top \begin{bmatrix} -2ZZ^\top & 0 \\ 0 & 2Z^\top Z \end{bmatrix} \Theta \quad (2.96)$$

holds as well.

We now proceed with μ_P^{QR} . Let $\varepsilon \in \mathbb{R}$ be a parameter and let

$$X(\varepsilon) := \Theta(I + \varepsilon[\xi, P])\Theta^\top = \begin{bmatrix} I_m & \varepsilon Z \\ -\varepsilon Z^\top & I_{n-m} \end{bmatrix} \quad (2.97)$$

with QR -factorisation

$$X(\varepsilon) = X(\varepsilon)_Q X(\varepsilon)_R, \quad (2.98)$$

i.e.,

$$X(\varepsilon)_Q = \begin{bmatrix} R_{11}^{-1}(\varepsilon) & \varepsilon Z R_{22}^{-1}(\varepsilon) \\ -\varepsilon Z^\top R_{11}^{-1}(\varepsilon) & R_{22}^{-1}(\varepsilon) \end{bmatrix}, \quad (2.99)$$

where the Cholesky factors R_{ii} are defined via

$$\begin{aligned} R_{11}^\top(\varepsilon) R_{11}(\varepsilon) &= I_m + \varepsilon^2 Z Z^\top, \\ R_{22}^\top(\varepsilon) R_{22}(\varepsilon) &= I_{n-m} + \varepsilon^2 Z^\top Z. \end{aligned} \quad (2.100)$$

Obviously,

$$R_{11}(0) = I_m \quad \text{and} \quad R_{22}(0) = I_{n-m}. \quad (2.101)$$

Therefore, taking the first order derivatives in (2.100) and evaluating at $\varepsilon = 0$ gives

$$\begin{aligned}\dot{R}_{11}^\top(0) + \dot{R}_{11}(0) &= 0, \\ \dot{R}_{22}^\top(0) + \dot{R}_{22}(0) &= 0,\end{aligned}\tag{2.102}$$

which imply $\dot{R}_{11}(0) = 0$ and $\dot{R}_{22}(0) = 0$. Furthermore, taking second order derivatives at $\varepsilon = 0$ and using (2.101) and (2.102) gives

$$\begin{aligned}\ddot{R}_{11}^\top(0) + \ddot{R}_{11}(0) &= 2ZZ^\top, \\ \ddot{R}_{22}^\top(0) + \ddot{R}_{22}(0) &= 2Z^\top Z.\end{aligned}\tag{2.103}$$

Using (2.101), (2.102) and (2.103) we compute the derivatives of the inverses as

$$\begin{aligned}\left.\frac{d}{d\varepsilon}R_{11}^{-1}(\varepsilon)\right|_{\varepsilon=0} &= 0, \\ \left.\frac{d}{d\varepsilon}R_{22}^{-1}(\varepsilon)\right|_{\varepsilon=0} &= 0\end{aligned}\tag{2.104}$$

and

$$\begin{aligned}\left.\frac{d^2}{d\varepsilon^2}R_{11}^{-1}(\varepsilon)\right|_{\varepsilon=0} &= -\ddot{R}_{11}(0), \\ \left.\frac{d^2}{d\varepsilon^2}R_{22}^{-1}(\varepsilon)\right|_{\varepsilon=0} &= -\ddot{R}_{22}(0).\end{aligned}\tag{2.105}$$

Therefore,

$$\begin{aligned}X(0)_Q &= I, \\ \left.\frac{d}{d\varepsilon}X(\varepsilon)_Q\right|_{\varepsilon=0} &= \begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix}, \\ \left.\frac{d^2}{d\varepsilon^2}X(\varepsilon)_Q\right|_{\varepsilon=0} &= \begin{bmatrix} -\ddot{R}_{11}(0) & 0 \\ 0 & -\ddot{R}_{22}(0) \end{bmatrix}\end{aligned}\tag{2.106}$$

and finally,

$$\begin{aligned}\left.\frac{d^2}{d\varepsilon^2}\mu_P^{\text{QR}}(\varepsilon\xi)\right|_{\varepsilon=0} &= \Theta^\top \left(\dot{X}(0)_Q \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \dot{X}^\top(0)_Q + 2\dot{X}(0)_Q \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \dot{X}^\top(0)_Q \right) \Theta \\ &= \Theta^\top \begin{bmatrix} -\ddot{R}_{11}(0) - \ddot{R}_{11}^\top(0) & 0 \\ 0 & 2Z^\top Z \end{bmatrix} \Theta \\ &= \Theta^\top \begin{bmatrix} -2ZZ^\top & 0 \\ 0 & 2Z^\top Z \end{bmatrix} \Theta\end{aligned}\tag{2.107}$$

as required. \square

2.2 Gradients and Hessians

Let $F : \text{Sym}_n \rightarrow \mathbb{R}$ be a smooth function and let $f := F|_{\text{Gr}_{m,n}}$ denote its restriction to the Grassmannian. Let $\nabla_F(P) \in \text{Sym}_n$ be the gradient of F in Sym_n evaluated at P . Let $\text{H}_F(P) : \text{Sym}_n \times \text{Sym}_n \rightarrow \mathbb{R}$ denote the Hessian form of F evaluated at P . We also consider $\text{Hess}_F(P) : \text{Sym}_n \rightarrow \text{Sym}_n$ as the corresponding linear map. Gradient and Hessian are formed using the Euclidean (Frobenius) inner product on Sym_n .

The next result computes the Riemannian gradient and Riemannian Hessian of the restriction f with respect to the induced Euclidean Riemannian metric on the Grassmannian (and thus also for the normal Riemannian metric).

Theorem 2.4 *Let $f : \text{Gr}_{m,n} \rightarrow \mathbb{R}$. The Riemannian gradient, grad_f , and the Riemannian Hessian operator $\text{Hess}_f(P) : T_P \text{Gr}_{m,n} \rightarrow T_P \text{Gr}_{m,n}$ are given as*

$$\text{grad}_f(P) = \text{ad}_P^2(\nabla_F(P)) = [P, [P, \nabla_F(P)]], \quad (2.108)$$

and

$$\text{Hess}_f(P)(\xi) = \text{ad}_P^2(\text{Hess}_F(P)(\xi)) - \text{ad}_P \text{ad}_{\nabla_F(P)} \xi, \quad (2.109)$$

for all $\xi \in T_P \text{Gr}_{m,n}$.

Proof The first part follows immediately from the well known fact, that the Riemannian gradient of f coincides with the orthogonal projection of ∇_F onto the tangent space $T_P \text{Gr}_{m,n}$. Since the orthogonal projection operator onto the tangent space is given by ad_P^2 , this proves the first claim.

For the second part, consider a geodesic curve $P(t)$ with $\dot{P}(0) = \xi$. Then $\ddot{P} = -[\dot{P}, [\dot{P}, P]]$ and therefore the Riemannian Hessian form is

$$\begin{aligned} \text{H}_f(P(0))(\xi, \xi) &:= \left. \frac{d^2(F \circ P)(t)}{dt^2} \right|_{t=0} \\ &= \text{H}_F(P(0))(\xi, \xi) + D F(P(0)) \ddot{P}(0) \\ &= \text{H}_F(P(0))(\xi, \xi) - D F(P(0)) [\xi, [\xi, P(0)]] \\ &= \text{H}_F(P(0))(\xi, \xi) - \text{tr} \left(\nabla_F(P(0)) [\xi, [\xi, P(0)]] \right). \end{aligned} \quad (2.110)$$

Thus by polarization

$$\begin{aligned} \text{H}_{f(P)}(\xi, \eta) &= \text{H}_F(P)(\xi, \eta) - \frac{1}{2} \text{tr}(\nabla_F(P) [\xi, [\eta, P]]) - \frac{1}{2} \text{tr}(\nabla_F(P) [\eta, [\xi, P]]) \\ &= \text{tr} \left((\text{Hess}_F(P)(\xi) - \frac{1}{2} [P, [\nabla_F(P), \xi]] - \frac{1}{2} [[\xi, P], \nabla_F(P)]) \eta \right) \\ &= \text{tr} \left((\text{Hess}_F(P)(\xi) - \frac{1}{2} [P, [\nabla_F(P), \xi]] - \frac{1}{2} [\nabla_F(P), [P, \xi]]) \eta \right) \\ &= \text{tr} \left(\text{ad}_P^2 \left(\text{Hess}_F(P)(\xi) - \frac{1}{2} [P, [\nabla_F(P), \xi]] - \frac{1}{2} [\nabla_F(P), [P, \xi]] \right) \eta \right). \end{aligned} \quad (2.111)$$

This implies that the Riemannian Hessian operator is given as

$$\begin{aligned}
\mathfrak{Hess}_f(P)(\xi) &= \text{ad}_P^2 \left(\text{Hess}_F(P)(\xi) - \frac{1}{2} \text{ad}_P \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_{\nabla_F(P)} \text{ad}_P \xi \right) \\
&= \text{ad}_P^2 \left(\text{Hess}_F(P)(\xi) \right) - \frac{1}{2} \text{ad}_P^3 \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_P^2 \text{ad}_{\nabla_F(P)} \text{ad}_P \xi \\
&= \text{ad}_P^2 \left(\text{Hess}_F(P)(\xi) \right) - \text{ad}_P^3 \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_P^2 [\text{ad}_{\nabla_F(P)}, \text{ad}_P] \xi \\
&= \text{ad}_P^2 \left(\text{Hess}_F(P)(\xi) \right) - \text{ad}_P \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_P^2 [\text{ad}_{\nabla_F(P)}, \text{ad}_P] \xi \\
&= \text{ad}_P^2 \left(\text{Hess}_F(P)(\xi) \right) - \text{ad}_P \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_P^2 \text{ad}_{[\nabla_F(P), P]} \xi.
\end{aligned} \tag{2.112}$$

The result thus follows from the following lemma. \square

Lemma 2.2 *For any tangent vector $\xi \in T_P \text{Gr}_{m,n}$ and any $A \in \text{Sym}_n$ one has*

$$\text{ad}_P \text{ad}_{[A, P]} \xi = 0. \tag{2.113}$$

Proof Without loss of generality we can assume that

$$P = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}, \quad \xi = \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix}, \quad A = \begin{bmatrix} A_1 & A_2 \\ A_2^\top & A_3 \end{bmatrix}. \tag{2.114}$$

Then

$$[A, P] = \begin{bmatrix} 0 & -A_2 \\ A_2^\top & 0 \end{bmatrix} \tag{2.115}$$

and

$$\text{ad}_{[A, P]} \xi = \begin{bmatrix} -ZA_2^\top - A_2Z^\top & 0 \\ 0 & A_2^\top Z + Z^\top A_2 \end{bmatrix}, \tag{2.116}$$

from which $\text{ad}_P \text{ad}_{[A, P]} \xi = 0$ follows by a straightforward computation. \square

As a consequence we obtain the following formulas for the Riemannian gradient and Riemannian Hessian operator of the Rayleigh quotient function.

Corollary 2.2 *Let $A \in \text{Sym}_n$. The Riemannian gradient and Riemannian Hessian operator of the Rayleigh quotient function*

$$f : \text{Gr}_{m,n} \rightarrow \mathbb{R}, \quad f(P) := \text{tr}(AP) \tag{2.117}$$

are

$$\begin{aligned}
\text{grad}_f(P) &= [P, [P, A]], \\
\mathfrak{Hess}_f(P) &= -\text{ad}_P \circ \text{ad}_A,
\end{aligned} \tag{2.118}$$

respectively.

As another example, let us consider the function

$$\begin{aligned} F : \text{Sym}_n &\rightarrow \mathbb{R}, \\ P &\mapsto \|(I - P)AP\|^2 = \text{tr}(I - P)APA^\top \end{aligned} \quad (2.119)$$

for an arbitrary, not necessarily symmetric matrix $A \in \mathbb{R}^{n \times n}$. Note that the global minima of the restriction $f := F|_{\text{Gr}_{m,n}}$ to the Graßmannian are exactly the projection operators corresponding to the m -dimensional invariant subspaces of A . The gradient and Hessian operator on Sym_n are computed as:

$$\begin{aligned} \left. \frac{d}{d\varepsilon} F(P + \varepsilon H) \right|_{\varepsilon=0} &= \text{tr}(-HAPA^\top + (I - P)AHA^\top) \\ &= \text{tr} H(A^\top(I - P)A - APA^\top), \\ \left. \frac{d^2}{d\varepsilon^2} F(P + \varepsilon H) \right|_{\varepsilon=0} &= -2 \text{tr} HAH A^\top, \end{aligned} \quad (2.120)$$

by polarisation for $H, K \in \text{Sym}_n$

$$\frac{1}{4}(-2 \text{tr}(H + K)A(H + K)A^\top + 2(H - K)A(H - K)A^\top) = -\text{tr} H(AKA^\top + A^\top KA)$$

consequently,

$$\begin{aligned} \nabla_F(P) &= A^\top(I - P)A - APA^\top, \\ \text{Hess}_F(P)(\xi) &= -A^\top \xi A - A\xi A^\top. \end{aligned} \quad (2.121)$$

This leads to the following explicit description of the Riemannian gradient and Riemannian Hessian operators on the Graßmannian.

Corollary 2.3 *Let $A \in \mathbb{R}^{n \times n}$. The Riemannian gradient and Riemannian Hessian operator of*

$$f : \text{Gr}_{m,n} \rightarrow \mathbb{R}, \quad f(P) := \|(I - P)AP\|^2 \quad (2.122)$$

are

$$\begin{aligned} \text{grad}_f(P) &= [P, [P, A^\top A - A^\top P A - APA^\top]], \\ \text{Hess}_f(P)(\xi) &= -[P, [P, A^\top \xi A + A\xi A^\top]] - [P, [A^\top A - A^\top P A - APA^\top, \xi]], \end{aligned} \quad (2.123)$$

respectively.

3 Geometry of the Lagrange Graßmannian

In this section we develop an analogous theory for the manifold of Lagrangian subspaces in \mathbb{R}^{2n} . Thus we consider the *Lagrange Graßmann manifold* LG_n , consisting of all n -dimensional Lagrangian subspaces of \mathbb{R}^{2n} with respect to the standard symplectic form

$$J := \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}. \quad (3.1)$$

Recall, that an n -dimensional subspace $V \subset \mathbb{R}^{2n}$ is called Lagrangian, if

$$v^\top J v = 0 \quad (3.2)$$

for all $v \in V$. Instead of interpreting the elements of the Lagrange Graßmann manifold as maximal isotropic subspaces, we prefer to view them in an equivalent way as a certain subclass of symmetric projection operators. Note that, if P is the symmetric projection operator onto an n -dimensional linear subspace V , then the condition $PJP = 0$ is equivalent to V being Lagrangian. Thus we define the *Lagrange Graßmannian*

$$\text{LG}_n := \{P \in \text{Sym}_{2n} \mid P^2 = P, \text{tr } P = n, PJP = 0\} \quad (3.3)$$

as the manifold of rank n symmetric projection operators of \mathbb{R}^{2n} , satisfying the Lagrangian subspace condition $PJP = 0$. Note, that LG_n is a compact, connected submanifold of the Graßmannian $\text{Gr}_{n,2n}$. In order to obtain a deeper understanding of the geometry of this set, we observe that LG_n is a homogeneous space for the action of the orthogonal symplectic group. Let

$$\text{GL}_n := \{X \in \mathbb{R}^{n \times n} \mid \det X \neq 0\} \quad (3.4)$$

and let

$$\text{OSp}_{2n} := \{T \in \text{GL}_{2n} \mid T^\top J T = J, T \in \text{SO}_{2n}\} \quad (3.5)$$

denote the Lie group of orthogonal symplectic transformations. Let

$$\mathfrak{osp}_{2n} = \{X \in \mathfrak{so}_{2n} \mid X^\top J + JX = 0\} \quad (3.6)$$

denote the associated Lie algebra of skew-symmetric Hamiltonian $2n \times 2n$ -matrices. Thus the elements of \mathfrak{osp}_{2n} are exactly the real $2n \times 2n$ -matrices T of the form

$$T := \begin{bmatrix} A & -B \\ B & A \end{bmatrix} \quad (3.7)$$

defined by the condition, that $A + \imath B \in \mathfrak{u}_n$, i.e. $A + \imath B$ is skew-Hermitian, i.e., $A \in \mathfrak{so}_n$ and $B \in \text{Sym}_n$, where

$$\mathfrak{u}_n := \{X \in \mathbb{C}^{n \times n} \mid X^* = -X\} \quad (3.8)$$

and the asterisk symbol denotes complex conjugate transpose and $\imath := \sqrt{-1}$. Similarly, the elements of OSp_{2n} are the real $2n \times 2n$ -matrices ξ of the form

$$\xi := e \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \quad \text{satisfying} \quad X \in \mathfrak{so}_n, Y \in \text{Sym}_n. \quad (3.9)$$

In particular, OSp_{2n} is isomorphic to the unitary group

$$\text{U}_n := \{X \in \mathbb{C}^{n \times n} \mid X^* X = I_n\}. \quad (3.10)$$

The orthogonal symplectic group OSp_{2n} acts transitively on LG_n via

$$\begin{aligned} \sigma : \text{OSp}_{2n} \times \text{LG}_n &\rightarrow \text{LG}_n, \\ (T, P) &\mapsto T^\top P T, \end{aligned} \quad (3.11)$$

with the stabilizer subgroup of

$$P := \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \in \text{LG}_n \quad (3.12)$$

given as the set of all block-diagonal matrices

$$T = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \quad \text{with} \quad A \in \text{O}_n. \quad (3.13)$$

Therefore LG_n is a homogeneous space that can be identified with U_n/O_n .

Theorem 3.1 (a) *The Lagrange Graßmannian LG_n is a smooth, compact submanifold of Sym_{2n} of dimension $\frac{n(n+1)}{2}$.*
 (b) *The tangent space of LG_n at an element $P \in \text{LG}_n$ is given as*

$$T_P \text{LG}_n = \{[P, \Omega] \mid \Omega \in \mathfrak{osp}_{2n}\}. \quad (3.14)$$

Since the tangent space $T_P \text{LG}_n \subset \text{Sym}_{2n}$ is a subset of Sym_{2n} , we can define the normal space at P to be the vector space

$$N_P \text{LG}_n = (T_P \text{LG}_n)^\perp := \{X \in \text{Sym}_{2n} \mid \text{tr}(XY) = 0 \text{ for all } Y \in T_P \text{LG}_n\}. \quad (3.15)$$

Proposition 3.1 *Let $P \in \text{LG}_n$ be arbitrary.*

1. *The normal subspace in Sym_{2n} at P is given as*

$$N_P \text{LG}_n = \left\{ X - \frac{1}{2} \text{ad}_P^2(JXJ + X) \mid X \in \text{Sym}_{2n} \right\}. \quad (3.16)$$

2. *The linear map*

$$\begin{aligned} \pi : \text{Sym}_{2n} &\rightarrow \text{Sym}_{2n}, \\ X &\mapsto \frac{1}{2}[P, [P, JXJ + X]] \end{aligned} \quad (3.17)$$

is a self-adjoint projection operator onto $T_P \text{LG}_n$ with kernel $N_P \text{LG}_n$.

Proof To prove the first statement let $\Omega \in \mathfrak{osp}_{2n}$, $X \in \text{Sym}_{2n}$ and $P \in \text{LG}_n$ be arbitrary. Then for $[P, \Omega] \in T_P \text{LG}_n$ and $X - \frac{1}{2} \text{ad}_P^2(JXJ + X) \in N_P \text{LG}_n$ we have

$$\begin{aligned} \text{tr} \left([P, \Omega] \left(X - \frac{1}{2} \text{ad}_P^2(JXJ + X) \right) \right) &= \text{tr} \left(\Omega \left(\frac{1}{2} \text{ad}_P^3(JXJ + X) - \text{ad}_P X \right) \right) \\ &= \text{tr} \left(\Omega \left(\frac{1}{2} \text{ad}_P(JXJ + X) - \text{ad}_P X \right) \right) \\ &= \frac{1}{2} \text{tr} (\Omega \text{ad}_P(JXJ - X)) \\ &= \text{tr} (\Omega P(JXJ - X)) \\ &= \text{tr} (X \Omega(JPJ - P)) \\ &= 0, \end{aligned} \quad (3.18)$$

where we have used Lemma 2.1, Ω being skew-symmetric and Hamiltonian, and the easily verified identity

$$JPJ - P = -I_{2n}, \text{ for all } P \in \text{LG}_n. \quad (3.19)$$

By (3.18), $T_P \text{LG}_n$ and $N_P \text{LG}_n$ are orthogonal subspaces of Sym_{2n} with respect to the Frobenius inner product. Analogously to Proposition 2.1 we now see that $\text{Sym}_{2n} = T_P \text{LG}_n \oplus N_P \text{LG}_n$ holds true as well:

Every $P \in \text{LG}_n$ can be written as

$$P = Q^\top \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} Q \quad (3.20)$$

for some $Q \in \text{OSp}_{2n}$. Note that for all $\Omega \in \mathfrak{osp}_{2n}$, and for all $X, S \in \text{Sym}_{2n}$

$$\begin{aligned} \text{tr}(S[P, \Omega]) &= \text{tr}\left(QSQ^\top \left[\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}, Q\Omega Q^\top\right]\right), \\ \text{tr}\left(S\left(X - \frac{1}{2}\text{ad}_P^2(JXJ + X)\right)\right) &= \text{tr}\left(QSQ^\top\left(QXQ^\top - \frac{1}{2}\text{ad}_{\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}}^2(JQXQ^\top + QXQ^\top)\right)\right). \end{aligned} \quad (3.21)$$

By (3.21) and $\text{Sym}_{2n} \rightarrow Q(\text{Sym}_{2n})Q^\top$ being an isomorphism, without loss of generality we might assume that

$$P = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}. \quad (3.22)$$

Assume there exists an $S \in \text{Sym}_{2n}$ being orthogonal to both subspaces. We will show the implication

$$\left. \begin{aligned} \text{tr}(S \text{ad}_P \Omega) &= 0 \quad \text{for all } \Omega \in \mathfrak{osp}_{2n} \\ \text{tr}\left(S\left(X - \frac{1}{2}\text{ad}_P^2(JXJ + X)\right)\right) &= 0 \quad \text{for all } X \in \text{Sym}_{2n} \end{aligned} \right\} \implies S = 0. \quad (3.23)$$

Partition $S \in \text{Sym}_{2n}$ as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^\top & S_{22} \end{bmatrix} \quad (3.24)$$

then

$$\text{tr}(S \text{ad}_P \Omega) = 0 \quad \text{for all } \Omega \in \mathfrak{osp}_{2n} \iff S_{12} \in \mathfrak{so}_n, \quad (3.25)$$

where we have used the symmetry of the (12)–block of Ω , see (3.6) and (3.7). Moreover,

$$\begin{aligned} \text{tr}\left(S\left(X - \frac{1}{2}\text{ad}_P^2(JXJ + X)\right)\right) &= \text{tr}\left(SX - \frac{1}{2}(\text{ad}_P^2 S)(JXJ + X)\right) \\ &= \text{tr}\left(\begin{bmatrix} S_{11} & \frac{S_{12} - S_{12}^\top}{2} \\ \frac{S_{12}^\top - S_{12}}{2} & S_{22} \end{bmatrix} X\right). \end{aligned} \quad (3.26)$$

Therefore,

$$\text{tr}\left(S\left(X - \frac{1}{2}\text{ad}_P^2(JXJ + X)\right)\right) = 0 \text{ for all } X \in \text{Sym}_{2n} \iff \begin{cases} S_{12} \in \text{Sym}_n, \\ S_{11} = S_{22} = 0. \end{cases} \quad (3.27)$$

Together with (3.25) we conclude $S = 0$, i.e., $\text{Sym}_{2n} = T_P \text{LG}_n \oplus N_P \text{LG}_n$ as required.

Now we prove the second claim. By the same reasoning as above we again might assume that

$$P = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}. \quad (3.28)$$

Let

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^\top & X_{22} \end{bmatrix}. \quad (3.29)$$

Since

$$\pi(X) = \frac{1}{2}\text{ad}_P^2(JXJ + X) = \begin{bmatrix} 0 & \frac{X_{12} + X_{12}^\top}{2} \\ \frac{X_{12} + X_{12}^\top}{2} & 0 \end{bmatrix}, \quad (3.30)$$

we see that $\pi^2(X) = \pi(X)$ and moreover $\text{im } \pi = T_P \text{LG}_n$. For any $X - \frac{1}{2} \text{ad}_P^2(JXJ + X) \in N_P \text{LG}_n$ we have

$$\pi\left(X - \frac{1}{2} \text{ad}_P^2(JXJ + X)\right) = \pi(X) - \pi^2(X) = 0, \quad (3.31)$$

and by counting dimensions $\ker \pi = N_P \text{LG}_n$. Finally, for all $X, Y \in \text{Sym}_{2n}$ and by using (3.30)

$$\begin{aligned} \langle \pi(X), Y \rangle &= \text{tr}\left(\frac{1}{2} \text{ad}_P^2(JXJ + X)Y\right) \\ &= \frac{1}{2} \text{tr}\left(\begin{bmatrix} 0 & X_{12} + X_{12}^\top \\ X_{12} + X_{12}^\top & 0 \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{12}^\top & Y_{22} \end{bmatrix}\right) \\ &= \frac{1}{2} \text{tr}\left(\begin{bmatrix} X_{11} & X_{12} \\ X_{12}^\top & X_{22} \end{bmatrix} \begin{bmatrix} 0 & Y_{12} + Y_{12}^\top \\ Y_{12} + Y_{12}^\top & 0 \end{bmatrix}\right) \\ &= \text{tr}\left(X \frac{1}{2} \text{ad}_P^2(JYJ + Y)\right) \\ &= \langle X, \pi(Y) \rangle. \end{aligned} \quad (3.32)$$

Thus π is self-adjoint and the result follows. \square

Fortunately, the discussion of Riemannian metrics carries directly over from the Graßmannian case to the case of the Lagrange Graßmannian. We therefore omit the proof.

Consider the surjective linear map

$$\begin{aligned} \text{ad}_P : \mathfrak{osp}_{2n} &\rightarrow T_P \text{LG}_n, \\ \Omega &\mapsto [P, \Omega] \end{aligned} \quad (3.33)$$

with kernel

$$\ker \text{ad}_P = \{\Omega \in \mathfrak{osp}_{2n} \mid P\Omega = \Omega P\}. \quad (3.34)$$

We regard \mathfrak{osp}_{2n} as an inner product space, endowed with the Frobenius inner product $\langle \Omega_1, \Omega_2 \rangle = \text{tr}(\Omega_1^\top \Omega_2)$. Then ad_P induces an isomorphism of vector spaces

$$\widehat{\text{ad}}_P : (\ker \text{ad}_P)^\perp \rightarrow T_P \text{LG}_n \quad (3.35)$$

and therefore induces an isometry of inner product spaces, by defining an inner product on $T_P \text{LG}_n$ via

$$\langle \langle X, Y \rangle \rangle_P := -\text{tr}(\widehat{\text{ad}}_P^{-1}(X) \widehat{\text{ad}}_P^{-1}(Y)) \quad (3.36)$$

called the normal Riemannian metric.

Proposition 3.2 *The Euclidean and normal Riemannian metrics on the Lagrange Graßmannian LG_n coincide, i.e. for all $P \in \text{LG}_n$ and for all $X, Y \in T_P \text{LG}_n$ we have*

$$\text{tr}(X^\top Y) = -\text{tr}\left(\widehat{\text{ad}}_P^{-1}(X) \widehat{\text{ad}}_P^{-1}(Y)\right). \quad (3.37)$$

Since a solution to (2.38) with an initial value $P_0 \in \text{LG}_n$ and $\dot{P}_0 \in T_{P_0} \text{LG}_n$ is fully contained in LG_n , and since geodesics are unique, the geodesics of LG_n are also described by that equation.

Theorem 3.2 *The geodesics of LG_n are exactly the solutions of the second order differential equation*

$$\ddot{P} + [\dot{P}, [\dot{P}, P]] = 0. \quad (3.38)$$

The unique geodesic $P(t)$ with initial conditions $P(0) = P_0 \in \text{LG}_n$, $\dot{P}(0) = \dot{P}_0 \in T_{P_0} \text{LG}_n$ is given by

$$P(t) = e^{t[\dot{P}_0, P_0]} P_0 e^{-t[\dot{P}_0, P_0]}. \quad (3.39)$$

We now consider *local parametrizations* for the Lagrange Graßmannian as well

$$\mu_P : T_P \text{LG}_n \rightarrow \text{LG}_n \quad (3.40)$$

satisfying

$$\mu_P(0) = P \text{ and } D\mu_P(0) = \text{id}. \quad (3.41)$$

3.1 Parametrizations and coordinates for the Lagrange Graßmannian

3.1.1 Riemannian normal coordinates

As before Riemannian normal coordinates are defined through

$$\begin{aligned} \exp_P : T_P \text{LG}_n &\rightarrow \text{LG}_n, \\ \exp_P(\xi) &= e^{[\xi, P]} P e^{-[\xi, P]}. \end{aligned} \quad (3.42)$$

Given any $\Theta \in \text{OSp}_{2n}$ with

$$P = \Theta^\top \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \Theta \quad (3.43)$$

and

$$[\xi, P] = \Theta^\top \begin{bmatrix} 0 & Z \\ -Z & 0 \end{bmatrix} \Theta, \quad Z \in \text{Sym}_n. \quad (3.44)$$

We obtain

$$\exp_P(\xi) = \Theta^\top \begin{bmatrix} \cos Z & \\ -\sin Z & \end{bmatrix} [\cos Z \quad -\sin Z] \Theta. \quad (3.45)$$

3.1.2 QR-coordinates

Let P and $[\xi, P]$ as in (3.43) and (3.44). We define smooth QR-coordinates by the map

$$\begin{aligned} \mu_P^{\text{QR}} : T_P \text{LG}_n &\rightarrow \text{LG}_n, \\ \xi &\mapsto (I + [\xi, P])_Q P (I + [\xi, P])_Q^\top. \end{aligned} \quad (3.46)$$

Analogous to Section 2.1.2 we define

$$\begin{bmatrix} I_n & Z \\ -Z & I_n \end{bmatrix}_Q := \begin{bmatrix} R^{-1} & ZR^{-1} \\ -ZR^{-1} & R^{-1} \end{bmatrix} \quad (3.47)$$

where R denotes the unique Cholesky factor that solves $R^\top R = I + Z^2$. Note that the Q-factor in (3.47) is orthogonal and symplectic. The above map (3.46) is therefore well defined.

3.1.3 Cayley coordinates

Let P and $[\xi, P]$ be as in (3.43) and (3.44). For any skew-symmetric Hamiltonian matrix Ω the *Cayley transform*

$$\begin{aligned} \text{Cay} : \mathfrak{osp}_{2n} &\rightarrow \text{OSp}_{2n}, \\ \Omega &\rightarrow (2I + \Omega)(2I - \Omega)^{-1} \end{aligned} \quad (3.48)$$

is smooth and satisfies $\text{D Cay}(0) = \text{id}$. The Cayley coordinates are defined as

$$\begin{aligned} \mu_P^{\text{Cay}} : T_P \text{LG}_n &\rightarrow \text{LG}_n, \\ \xi &\mapsto \text{Cay}([\xi, P]) P \text{Cay}(-[\xi, P]). \end{aligned} \quad (3.49)$$

Therefore,

$$\mu_P^{\text{Cay}}(\xi) = \Theta^\top \begin{bmatrix} I_n - \frac{1}{4}Z^2 \\ -Z \end{bmatrix} \left(I_n + \frac{1}{4}Z^2 \right)^{-2} [I_n - \frac{1}{4}Z^2 - Z] \Theta. \quad (3.50)$$

3.2 Gradients and Hessians

Let $F : \text{Sym}_n \rightarrow \mathbb{R}$ be a smooth function and let $f := F|_{\text{LG}_n}$ denote its restriction to the Lagrange Graßmannian. Let $\nabla_F(P) \in \text{Sym}_n$ be the gradient of F in Sym_n evaluated at P . Let $\text{H}_F(P) : \text{Sym}_n \times \text{Sym}_n \rightarrow \mathbb{R}$ denote the Hessian form of F evaluated at P . We also consider $\text{Hess}_F(P) : \text{Sym}_n \rightarrow \text{Sym}_n$ as the corresponding linear map. Gradient and Hessian are formed using the Euclidean (Frobenius) inner product on Sym_n .

The next result computes the Riemannian gradient and Riemannian Hessian of the restriction f with respect to the induced Euclidean Riemannian metric on the Lagrange Graßmannian (and thus also for the normal Riemannian metric).

Theorem 3.3 *Let $f : \text{LG}_n \rightarrow \mathbb{R}$ and consider the orthogonal projection operator π as defined by (3.17). The Riemannian gradient, \mathfrak{grad}_f , and the Riemannian Hessian operator $\mathfrak{Hess}_f(P) : T_P \text{LG}_n \rightarrow T_P \text{LG}_n$ are given as*

$$\mathfrak{grad}_f(P) = \pi(\nabla_F(P)) = \frac{1}{2} \text{ad}_P^2 \left(J(\nabla_F(P))J + \nabla_F(P) \right), \quad (3.51)$$

and

$$\begin{aligned} \mathfrak{Hess}_f(P)(\xi) &= \frac{1}{2} \text{ad}_P^2 \left(J \left(\text{Hess}_F(P)(\xi) \right) J + \text{Hess}_F(P)(\xi) \right) \\ &\quad - \frac{1}{2} \text{ad}_P^2 \left(J \left(\text{ad}_P \text{ad}_{\nabla_F(P)} \xi \right) J + \text{ad}_P \text{ad}_{\nabla_F(P)} \xi \right), \end{aligned} \quad (3.52)$$

for all $\xi \in T_P \text{Gr}_{m,n}$.

Proof The first part follows again from the fact, that the Riemannian gradient of f coincides with the orthogonal projection of ∇_F onto the tangent space $T_P \text{LG}_n$.

Analogously to the proof of Theorem 2.4

$$\mathfrak{H}_f(P(0))(\xi, \xi) = \text{H}_F(P(0))(\xi, \xi) - \text{tr} \left(\nabla_F(P(0)) [\xi, [\xi, P(0)]] \right). \quad (3.53)$$

Thus by polarization

$$\begin{aligned}\mathfrak{H}_f(P)(\xi, \eta) &= \text{tr} \left((\text{Hess}_F(P)(\xi) - \frac{1}{2}[P, [\nabla_F(P), \xi]] - \frac{1}{2}[\nabla_F(P), [P, \xi]] \eta \right) \\ &= \text{tr} \left(\pi \left(\text{Hess}_F(P)(\xi) - \frac{1}{2}[P, [\nabla_F(P), \xi]] - \frac{1}{2}[\nabla_F(P), [P, \xi]] \right) \eta \right).\end{aligned}\quad (3.54)$$

This implies that the Riemannian Hessian operator is given as

$$\begin{aligned}\mathfrak{Hess}_f(P)(\xi) &= \pi \left(\text{Hess}_F(P)(\xi) - \frac{1}{2} \text{ad}_P \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_{\nabla_F(P)} \text{ad}_P \xi \right) \\ &= \pi \left(\text{Hess}_F(P)(\xi) - \text{ad}_P \text{ad}_{\nabla_F(P)} \xi - \frac{1}{2} \text{ad}_{[\nabla_F(P), P]} \xi \right) \\ &= \pi \left(\text{Hess}_F(P)(\xi) \right) - \pi \left(\text{ad}_P \text{ad}_{\nabla_F(P)} \xi \right) - \pi \left(\frac{1}{2} \text{ad}_{[\nabla_F(P), P]} \xi \right).\end{aligned}\quad (3.55)$$

Together with Lemma 2.2 the Lemma 3.1 below implies

$$\pi \left(\frac{1}{2} \text{ad}_{[\nabla_F(P), P]} \xi \right) = 0. \quad (3.56)$$

The result follows. \square

Lemma 3.1 *For any tangent vector $\xi \in T_P \text{LG}_n$ and any $A \in \text{Sym}_n$ one has*

$$\text{ad}_P(J(\text{ad}_{[A, P]} \xi)J) = 0. \quad (3.57)$$

Proof Without loss of generality we can assume that

$$P = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}, \quad \xi = \begin{bmatrix} 0 & Z \\ Z & 0 \end{bmatrix}, \quad Z = Z^\top, \quad A = \begin{bmatrix} A_1 & A_2 \\ A_2^\top & A_3 \end{bmatrix}. \quad (3.58)$$

Then

$$J(\text{ad}_{[A, P]} \xi)J = \begin{bmatrix} -A_2^\top Z - ZA_2 & 0 \\ 0 & ZA_2^\top + A_2Z \end{bmatrix}, \quad (3.59)$$

from which $\text{ad}_P(J(\text{ad}_{[A, P]} \xi)J) = 0$ follows by a straightforward computation. \square

As a consequence we obtain the following formulas for the Riemannian gradient and Riemannian Hessian operator of the Rayleigh quotient function used to compute the n -dimensional dominant eigenspace of a real symmetric Hamiltonian $(2n \times 2n)$ -matrix.

Consider the set of real symmetric Hamiltonian $(2n \times 2n)$ -matrices \mathfrak{p}_{2n}

$$\mathfrak{p}_{2n} := \left\{ H \in \text{Sym}_{2n} \mid H = \begin{bmatrix} S & T \\ T & -S \end{bmatrix}, \quad T, S \in \text{Sym}_n \right\} \quad (3.60)$$

Note that

$$JKJ = K \quad \text{for all } K \in \mathfrak{p}_{2n}. \quad (3.61)$$

Moreover, from the theory of Cartan decompositions, see e.g. [9], the following commutator relations are well known

$$[\mathfrak{p}_{2n}, \mathfrak{p}_{2n}] \subset \mathfrak{osp}_{2n}, \quad [\mathfrak{p}_{2n}, \mathfrak{osp}_{2n}] \subset \mathfrak{p}_{2n} \quad (3.62)$$

together with the isomorphisms

$$Q^\top(\mathfrak{osp}_{2n})Q \cong \mathfrak{osp}_{2n}, \quad Q^\top(\mathfrak{p}_{2n})Q \cong \mathfrak{p}_{2n} \quad \text{for all } Q \in \text{OSp}_{2n}. \quad (3.63)$$

Corollary 3.1 *Given $H \in \mathfrak{p}_{2n}$ Let*

$$F : \text{Sym}_{2n} \rightarrow \mathbb{R}, \quad P \mapsto \text{tr}(HP), \quad (3.64)$$

with restriction

$$f := F|_{\text{LG}_n}. \quad (3.65)$$

The Riemannian gradient and Riemannian Hessian operator are

$$\begin{aligned} \text{grad}_f(P) &= [P, [P, A]], \\ \text{Hess}_f(P) &= -\text{ad}_P \circ \text{ad}_A, \end{aligned} \quad (3.66)$$

respectively.

Proof Because the function F is linear the Euclidean gradient is simply

$$\nabla_F(P) = H, \quad (3.67)$$

and the Euclidean Hessian operator vanishes

$$\text{Hess}_F(P) = 0. \quad (3.68)$$

We therefore get for the Riemannian gradient using (3.61) and Theorem 3.3

$$\begin{aligned} \text{grad}_f(P) &= \frac{1}{2} \text{ad}_P^2 \left(J(\nabla_F(P))J + \nabla_F(P) \right) \\ &= \frac{1}{2} \text{ad}_P^2 \left(JHJ + H \right) \\ &= [P, [P, H]]. \end{aligned} \quad (3.69)$$

For the Riemannian Hessian operator we need some preparation. Let $\xi = [P, \Omega] \in T_P \text{LG}_n$ be arbitrary, i.e., $\Omega \in \mathfrak{osp}_{2n}$ is arbitrary. Then there exists a $Q \in \text{OSp}_{2n}$ such that

$$\xi = Q^\top \left[\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}, Q\Omega Q^\top \right] Q. \quad (3.70)$$

But the commutator in (3.70) is an element of \mathfrak{p}_{2n} and therefore by (3.63) the same holds true for ξ independent of P . Consequently, $\text{ad}_H \xi \in \mathfrak{osp}_{2n}$ and $\text{ad}_P \text{ad}_H \xi \in \mathfrak{p}_{2n}$, and finally $J(\text{ad}_P \text{ad}_H \xi)J = \text{ad}_P \text{ad}_H \xi$. The formula for the Riemannian Hessian operator is now easily verified

$$\begin{aligned} \text{Hess}_f(P)(\xi) &= -\frac{1}{2} \text{ad}_P^2 \left(J(\text{ad}_P \text{ad}_{\nabla_F(P)} \xi)J + \text{ad}_P \text{ad}_{\nabla_F(P)} \xi \right) \\ &= -\frac{1}{2} \text{ad}_P^2 \left(J(\text{ad}_P \text{ad}_H \xi)J + \text{ad}_P \text{ad}_H \xi \right) \\ &= -\text{ad}_P \text{ad}_H \xi. \end{aligned} \quad (3.71)$$

□

4 Newton's method

In the following we propose a class of Newton-like algorithms to compute a nondegenerate critical point of a smooth cost function $f : \text{Gr}_{m,n} \rightarrow \mathbb{R}$. Local quadratic convergence of the proposed algorithm will be established. Parts of this section are based on the conference paper [7].

4.1 The Euclidean case

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a smooth function and let $x^* \in \mathbb{R}^n$ be a nondegenerate critical point of f , i.e. the Hessian operator $\text{Hess}_f(x^*)$ is invertible. Newton's method for f is the iteration

$$x_0 \in \mathbb{R}^n, x_{k+1} = N_f(x_k) := x_k - (\text{Hess}_f(x_k))^{-1} \nabla f(x_k). \quad (4.1)$$

Note that the iteration (4.1) is only defined if $\text{Hess}_f(x_k)$ is invertible for all $k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. However, since f is smooth, there exists an open neighborhood of x^* in which the Hessian operator is invertible.

It is well known that the point sequence $\{x_k\}_{k \in \mathbb{N}_0}$ generated by (4.1) is defined and converges locally quadratically to x^* provided that x_0 is sufficiently close to x^* . For more information see e.g. [11].

4.2 The Grassmannian case

Let $\{\mu_P\}_{P \in \text{Gr}_{m,n}}$ be a family of local parametrizations of $\text{Gr}_{m,n}$. Let $P^* \in \text{Gr}_{m,n}$ be a nondegenerate critical point of the smooth function $f : \text{Gr}_{m,n} \rightarrow \mathbb{R}$. If there exists an open neighborhood $U \subset \text{Gr}_{m,n}$ of P^* and a smooth map

$$\mu : U \times \mathbb{R}^{m(n-m)} \longrightarrow \text{Gr}_{m,n}$$

such that $\mu(P, x) = \mu_P(x)$ for all $P \in U$ and $x \in \mathbb{R}^{m(n-m)}$ we will call $\{\mu_P\}_{P \in \text{Gr}_{m,n}}$ a *locally smooth family of parametrizations around P^** . Let $\{\mu_P\}_{P \in \text{Gr}_{m,n}}$ and $\{\nu_P\}_{P \in \text{Gr}_{m,n}}$ be two locally smooth families of parametrizations around P^* . Consider the following iteration on $\text{Gr}_{m,n}$

$$P_0 \in \text{Gr}_{m,n}, P_{k+1} = \nu_{P_k} \left(N_{f \circ \mu_{P_k}}(0) \right) \quad (4.2)$$

where $N_{f \circ \mu_P}$ is defined in (4.1). The following theorem is an adaptation from [7], where it is stated and proved for arbitrary smooth manifolds.

Theorem 4.1 *Under the condition*

$$D\mu_{P^*}(0) = D\nu_{P^*}(0) \quad (4.3)$$

there exists an open neighborhood $V \subset \text{Gr}_{m,n}$ of P^ such that the point sequence $\{P_k\}_{k \in \mathbb{N}_0}$ generated by (4.2) converges quadratically to P^* provided $P_0 \in V$.*

Proof Let $\mu, \nu : U \times \mathbb{R}^{m(n-m)} \rightarrow \text{Gr}_{m,n}$ be smooth and such that $\mu(P, x) = \mu_P(x)$ and $\nu(P, x) = \nu_P(x)$ for all $P \in U$ and $x \in \mathbb{R}^{m(n-m)}$, where U is a neighborhood of P^* .

The derivative of the algorithm map

$$\begin{aligned} s : \text{Gr}_{m,n} &\rightarrow \text{Gr}_{m,n}, \\ P &\mapsto \nu \left(P, -(\text{Hess}_{f \circ \mu}(P, 0))^{-1} \nabla_{f \circ \mu}(P, 0) \right) \end{aligned} \quad (4.4)$$

at P^* is the linear map

$$Ds(P^*) : T_{P^*} \text{Gr}_{m,n} \rightarrow T_{P^*} \text{Gr}_{m,n} \quad (4.5)$$

defined by

$$\begin{aligned} Ds(P^*)h &= D_1 \nu \left(P^*, -(\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \nabla_{f \circ \mu}(P^*, 0) \right) h \\ &+ D_2 \nu \left(P^*, -(\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \nabla_{f \circ \mu}(P^*, 0) \right) \\ &\cdot \left(-D_P \left((\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \nabla_{f \circ \mu}(P^*, 0) \right) h \right). \end{aligned} \quad (4.6)$$

Here $D_i(\cdot)h$ denotes the derivative of (\cdot) with respect to the i -th argument in direction h , whereas, by abuse of notation, D_P denotes the differential operator to compute the derivative with respect to the argument P .

The first summand on the right side in (4.6) is easily computed as

$$D_1 \nu \left(P^*, -(\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \nabla_{f \circ \mu}(P^*, 0) \right) h = D_1 \nu(P^*, 0)h = h, \quad (4.7)$$

which is true because P^* is a critical point of f and the gradient therefore vanishes.

The second summand in (4.6) consists of two terms due to the chain rule. We first compute the left term giving

$$D_2 \nu \left(P^*, -(\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \nabla_{f \circ \mu}(P^*, 0) \right) = D_2 \nu(P^*, 0) \quad (4.8)$$

because P^* is critical. The evaluation of the right term is more involved.

$$\begin{aligned} -D_P \left((\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \nabla_{f \circ \mu}(P^*, 0) \right) h &= \\ - \left(D_P \left((\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} \right) h \right) \underbrace{\nabla_{f \circ \mu}(P^*, 0)}_{=0} & \\ - (\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} D_P (\nabla_{f \circ \mu}(P^*, 0)) h & \\ = - (\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} D_P (\nabla_{f \circ \mu}(P^*, 0)) h. & \end{aligned} \quad (4.9)$$

By the definition of the gradient one has for any $x \in \mathbb{R}^{m(n-m)}$

$$\langle \nabla_{f \circ \mu}(P, 0), x \rangle = Df(P) \cdot D_2 \mu(P, 0) \cdot x \quad (4.10)$$

and therefore using the critical point condition and the definition of the Hessian operator in terms of second derivatives

$$\begin{aligned} \langle D_P (\nabla_{f \circ \mu}(P^*, 0)) h, x \rangle &= D^2 f(P^*) \cdot (D_2 \mu(P^*, 0) \cdot x, h) \\ &\quad + \underbrace{Df(P^*)}_{=0} D_1 (D_2 \mu(P^*, 0) \cdot x) h \\ &= D^2 f(P^*) \cdot (D_2 \mu(P^*, 0) \cdot x, h) \\ &= D^2 f(P^*) \cdot (D_2 \mu(P^*, 0) \cdot x, D_2 \mu(P^*, 0) (D_2 \mu(P^*, 0))^{-1} h) \\ &= \langle \text{Hess}_{f \circ \mu}(P^*, 0)(x), (D_2 \mu(P^*, 0))^{-1} h \rangle \\ &= \langle \text{Hess}_{f \circ \mu}(P^*, 0) (D_2 \mu(P^*, 0))^{-1} h, x \rangle \end{aligned} \quad (4.11)$$

We now can conclude

$$D_P (\nabla_{f \circ \mu}(P^*, 0)) h = \text{Hess}_{f \circ \mu}(P^*, 0) (D_2 \mu(P^*, 0))^{-1} h \quad (4.12)$$

which in turn implies that

$$(\text{Hess}_{f \circ \mu}(P^*, 0))^{-1} D_P (\nabla_{f \circ \mu}(P^*, 0)) h = (D_2 \mu(P^*, 0))^{-1} h. \quad (4.13)$$

Summarizing our computations we have shown that

$$\begin{aligned} D s(P^*) h &= h - D_2 \nu(P^*, 0) (D_2 \mu(P^*, 0))^{-1} h \\ &= 0. \end{aligned} \quad (4.14)$$

Consider now a local representation of s in coordinate charts around P^* and $s(P^*) = P^*$. Let $\|\cdot\|$ denote any norm in the local coordinate space. By abuse of notation we will still speak of s , P^* and so on in reference to their local coordinate representations. Using a Taylor expansion of s around P^* , there exists a neighborhood \overline{V}_{P^*} of P^* such that the estimate

$$\|s(P) - P^*\| \leq \sup_{Q \in \overline{V}_{P^*}} \|D^2 s(Q)\| \cdot \|P - P^*\|^2 \quad (4.15)$$

holds for all $P \in \overline{V}_{P^*}$. Therefore, the subset $U \subset \overline{V}_{P^*}$

$$U := \{P \in \overline{V}_{P^*} \mid \sup_{Q \in \overline{V}_{P^*}} \|D^2 s(Q)\| \cdot \|P - P^*\| < 1\}$$

is a neighborhood of P^* that is invariant under s , and hence remains invariant under the iterations of s . This completes the proof of local quadratic convergence of the algorithm. \square

A few remarks are in order. Geometrically, the iteration (4.2) does the following. The current iteration point P_k is pulled back to Euclidean space via the local parametrization μ_{P_k} around P_k . Then one Euclidean Newton step is performed for the function expressed in local coordinates, followed by a projection back onto the Grassmannian using the local parametrization ν_{P_k} around P_k .

For the special choice $\{\mu_p\}_{p \in M} = \{\nu_p\}_{p \in M}$ both Riemannian normal coordinates (cf. Section 2.1.1), our iteration (4.2) is precisely the so-called Newton method along geodesics of D. Gabay [5], more recently also referred to as the intrinsic Newton method. This follows from the lemma below.

Lemma 4.1 *Let $f : \text{Gr}_{m,n} \rightarrow \mathbb{R}$ be a smooth function. For all $P \in \text{Gr}_{m,n}$, $\xi \in T_P \text{Gr}_{m,n}$ and any $\text{type} \in \{\text{exp}, \text{QR}, \text{Cay}\}$ we have*

$$\nabla_{f \circ \mu_P^{\text{type}}}(0) = \text{grad}_f(P) \quad (4.16)$$

and

$$\text{Hess}_{f \circ \mu_P^{\text{type}}}(0)(\xi) = \mathfrak{Hess}_f(P)(\xi). \quad (4.17)$$

Proof By Remark 2.1 the unique geodesic through a point $P \in \text{Gr}_{m,n}$ in direction $\xi \in T_P \text{Gr}_{m,n}$ is given by $P(t) = \mu_P^{\text{exp}}(t\xi)$ and hence

$$\langle \langle \text{grad}_f(P), \xi \rangle \rangle_P = \left. \frac{d}{d\varepsilon} (f \circ \mu_P^{\text{exp}})(\varepsilon) \right|_{\varepsilon=0} = \langle \nabla_{f \circ \mu_P^{\text{exp}}}(0), \xi \rangle, \quad (4.18)$$

which by Proposition 2.3 implies (4.16) for $\text{type} = \text{exp}$. The result for $\text{type} = \text{QR}$ and $\text{type} = \text{Cay}$ then follows from (2.92). By the same line of arguments (4.17) follows from

$$\langle \langle \mathfrak{Hess}_f(P)(\xi), \xi \rangle \rangle_P = \left. \frac{d^2}{d\varepsilon^2} (f \circ \mu_P^{\text{exp}})(\varepsilon) \right|_{\varepsilon=0} = \langle \text{Hess}_{f \circ \mu_P^{\text{exp}}}(0)(\xi), \xi \rangle \quad (4.19)$$

and Theorem 2.3. \square

4.3 The Lagrange Graßmannian case

All the above results carry over literally to Newton-like algorithms on the Lagrange Graßmannian by substituting the respective formulas.

4.4 Algorithms

We conclude by presenting several specific instances of the resulting algorithms. We discuss the case of smooth functions $F : \text{Sym}_n \rightarrow \mathbb{R}$ with restriction $f := F|_{\text{Gr}_{m,n}}$ to the Graßmannian, as well as the special cases of the Rayleigh quotient function on the Graßmannian and the Lagrange Graßmannian. Furthermore, we consider the previously introduced nonlinear trace function for invariant subspace computations on the Graßmannian. In all cases we choose $\{\mu_P\}$ as the Riemannian normal coordinates and $\{\nu_P\}$ as the QR-coordinates, see Sections 2.1 and 3.1.

Recall that our convergence result requires the Hessian of the restricted function to be nondegenerate at the critical point.

We first formulate a preliminary form of the algorithm we are interested in.

Step 1.
Pick a rank m symmetric projection operator of \mathbb{R}^n , $P_0 \in \text{Gr}_{m,n}$, and set $j = 0$.

Step 2.
Solve

$$\text{ad}_{P_j}^2 \text{Hess}_F(P_j)(\text{ad}_{P_j} \Omega_j) - \text{ad}_{P_j} \text{ad}_{\nabla_F(P_j)} \text{ad}_{P_j} \Omega_j = -\text{ad}_{P_j}^2 \nabla_F(P_j)$$
for $\Omega_j \in \mathfrak{so}(n)$.

Step 3.
Solve

$$P_j = \Theta_j^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_j$$
for $\Theta_j \in \text{SO}_n$.

Step 4.
Compute

$$P_{j+1} = \Theta_j^\top \left(\Theta_j (I - \text{ad}_{P_j}^2 \Omega_j) \Theta_j^\top \right)_{\mathbb{Q}} \Theta_j P_j \Theta_j^\top \left(\Theta_j (I - \text{ad}_{P_j}^2 \Omega_j) \Theta_j^\top \right)_{\mathbb{Q}}^\top \Theta_j.$$

Step 5.
Set $j = j + 1$ and goto Step 2.

Here the expressions in Step 2 result from applying Lemma 4.1 and Theorem 2.4 and using the representation $\xi = [P, \Omega]$, $\Omega \in \mathfrak{so}(n)$ for an element $\xi \in T_P \text{Gr}_{m,n}$.

Inspecting the above algorithm, it is evident that it can be rewritten as an iteration in the $\Theta_j \in \text{SO}_n$ as follows.

Step 1.

Pick an orthogonal matrix $\Theta_0 \in \text{SO}_n$ corresponding to

$$P_0 = \Theta_0^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_0 \in \text{Gr}_{m,n},$$

and set $j = 0$.

Step 2.

Solve

$$\text{ad}_{P_j}^2 \text{Hess}_F(P_j)(\text{ad}_{P_j} \Omega_j) - \text{ad}_{P_j} \text{ad}_{\nabla_F(P_j)} \text{ad}_{P_j} \Omega_j = -\text{ad}_{P_j}^2 \nabla_F(P_j)$$

for $\Omega_j \in \mathfrak{so}(n)$.

Step 3.

Compute

$$\Theta_{j+1}^\top = \Theta_j^\top \left(\Theta_j (I - \text{ad}_{P_j}^2 \Omega_j) \Theta_j^\top \right)_\mathbb{Q} \text{ and } P_{j+1} = \Theta_{j+1}^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_{j+1}$$

Step 4.

Set $j = j + 1$ and goto Step 2.

Note that by equation (2.80) the term

$$\Theta_j (I - \text{ad}_{P_j}^2 \Omega_j) \Theta_j^\top = \begin{bmatrix} I_m & Z_j \\ -Z_j^\top & I_{n-m} \end{bmatrix} \quad (4.20)$$

of which we have to compute a QR-factorization in Step 3 has a nice block structure that can be exploited to get an efficient implementation. Locally quadratic convergence is guaranteed as long as the specific QR-factorization used is differentiable, cf. Section 2.1.2.

For specific functions the necessary computations might drastically simplify, as the following two examples show for the Rayleigh quotient function $F : \text{Sym}_n \rightarrow \mathbb{R}$, $F(P) = \text{tr}(AP)$, $A \in \text{Sym}_n$, cf. Corollary 2.2.

4.4.1 Rayleigh quotient on the Grassmannian

The equation we have to solve for $\Omega_j \in \mathfrak{so}(n)$ in Step 2 becomes

$$-\text{ad}_{P_j} \text{ad}_A \text{ad}_{P_j} \Omega_j = -\text{ad}_{P_j}^2 A \quad (4.21)$$

which is equivalent to

$$\Theta_j (\text{ad}_{P_j} \text{ad}_A \text{ad}_{P_j} \Omega_j) \Theta_j^\top = \Theta_j (\text{ad}_{P_j}^2 A) \Theta_j^\top \quad (4.22)$$

and, using

$$P_j = \Theta_j^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_j, \quad (4.23)$$

is equivalent to solving

$$\text{ad} \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \text{ad}_{\Theta_j A \Theta_j^\top} \text{ad} \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & Z_j \\ -Z_j^\top & 0 \end{bmatrix} = \text{ad}^2 \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} (\Theta_j A \Theta_j^\top) \quad (4.24)$$

for $Z_j \in \mathbb{R}^{m \times (n-m)}$. Denoting

$$\Theta_j A \Theta_j^\top = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{bmatrix} \quad (4.25)$$

we actually have to solve the Sylvester equation

$$A_{11}Z_j - Z_j A_{22} = A_{12}. \quad (4.26)$$

The resulting algorithm is exactly the algorithm presented in [7].

Algorithm 1: Rayleigh quotient on the Graßmannian.

Step 1.

Pick an orthogonal matrix $\Theta_0 \in \text{SO}_n$ corresponding to

$$P_0 = \Theta_0^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_0 \in \text{Gr}_{m,n},$$

and set $j = 0$.

Step 2.

Compute

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{bmatrix} = \Theta_j A \Theta_j^\top.$$

Step 3.

Solve the Sylvester equation

$$A_{11}Z_j - Z_j A_{22} = A_{12}.$$

for $Z_j \in \mathbb{R}^{m \times (n-m)}$.

Step 4.

Compute

$$\Theta_{j+1}^\top = \Theta_j^\top \begin{bmatrix} I_m & Z_j \\ -Z_j^\top & I_{n-m} \end{bmatrix}_Q \quad \text{and} \quad P_{j+1} = \Theta_{j+1}^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_{j+1}$$

Step 4.

Set $j = j + 1$ and goto Step 2.

Since the global maximum of the Rayleigh quotient function on the Graßmannian is a nondegenerate critical point, provided that there is a spectral gap after the m th largest eigenvalue of A , we immediately get the following result.

Theorem 4.2 *For almost all matrices $A \in \text{Sym}_n$ Algorithm 1 converges locally quadratically to the projector onto the m -dimensional dominant eigenspace.*

4.4.2 Rayleigh quotient on the Lagrange Graßmannian

The corresponding problem of optimizing the Rayleigh quotient function over the Lagrange Graßmann manifold can be treated completely analogous to the approach above. Thus let A denote a real symmetric Hamiltonian matrix of size $2n \times 2n$. The Newton algorithm for optimizing the trace function $\text{tr}(AP)$ over LG_n then is as follows.

Algorithm 2: Rayleigh quotient on the Lagrange Graßmannian.

Step 1.

Pick an orthogonal matrix $\Theta_0 \in \text{SO}_{2n}$ corresponding to

$$P_0 = \Theta_0^\top \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \Theta_0 \in \text{LG}_n,$$

and set $j = 0$.

Step 2.

Compute

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{12} & -A_{11} \end{bmatrix} = \Theta_j A \Theta_j^\top.$$

Step 3.

Solve the Lyapunov equation

$$A_{11} Z_j + Z_j A_{11} = A_{12}.$$

for the symmetric matrix $Z_j \in \text{Sym}_n$.

Step 4.

Compute

$$\Theta_{j+1}^\top = \Theta_j^\top \begin{bmatrix} I_n & Z_j \\ -Z_j & I_n \end{bmatrix}_Q \quad \text{and} \quad P_{j+1} = \Theta_{j+1}^\top \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \Theta_{j+1}$$

Step 4.

Set $j = j + 1$ and goto Step 2.

Algorithm 2 is almost identical to Algorithm 1 on the Graßmannian, except for the simpler Sylvester equation that is indeed a Lyapunov equation here. Again, we immediately get the following result.

Theorem 4.3 *Algorithm 2 converges locally quadratically to any nondegenerate critical point of the Rayleigh quotient function on the Lagrange Graßmannian.*

4.4.3 Invariant subspace computation

We now turn to the more complicated task of solving the optimization problem of the nonlinear trace function $\text{tr}((I - P)APA^\top)$ over the Graßmann manifold $\text{Gr}_{m,n}$. Here A denotes an arbitrary real $n \times n$ matrix. This is interesting as it leads to a locally quadratically convergent algorithm by solving only linear matrix equations.

Our method requires only orthogonal matrix calculations and a linear matrix solver. We omit the straightforward calculations that allow one to compute the Newton step in terms of the linear matrix equation appearing in Step 3 of the algorithm.

Algorithm 3: Invariant subspace function on the Graßmannian

Step 1.

Pick an orthogonal matrix $\Theta_0 \in \text{SO}_n$ corresponding to

$$P_0 = \Theta_0^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_0 \in \text{Gr}_{m,n},$$

and set $j = 0$.

Step 2.

Compute

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \Theta_j A \Theta_j^\top.$$

Step 3.

Solve the linear matrix equation

$$\begin{aligned} & A_{11}(A_{11}^\top Z_j - Z_j A_{22}^\top) - (A_{11}^\top Z_j - Z_j A_{22}^\top) A_{22} \\ & - A_{21}(Z_j^\top A_{12} + A_{21} Z_j) - (A_{12} Z_j^\top + Z_j A_{21}) A_{21}^\top = A_{21}^\top A_{22} - A_{11} A_{21}^\top. \end{aligned}$$

for $Z_j \in \mathbb{R}^{m \times (n-m)}$.

Step 4.

Compute

$$\Theta_{j+1}^\top = \Theta_j^\top \begin{bmatrix} I_m & Z_j \\ -Z_j^\top & I_{n-m} \end{bmatrix}_Q \quad \text{and} \quad P_{j+1} = \Theta_{j+1}^\top \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \Theta_{j+1}$$

Step 4.

Set $j = j + 1$ and goto Step 2.

We do not address the interesting but complicated issue how to solve the above linear matrix equation. Obviously one can always rewrite it as a linear equation on $\mathbb{R}^{m(n-m)}$ and then solve this, using matrix Kronecker products and vec-operations, by any linear equation solver. An alternative approach is to rewrite the equation in recursive form as

$$\begin{aligned} A_{11} X_j - X_j A_{22} &= A_{21}^\top (Z_{j-1}^\top A_{12} + A_{21} Z_{j-1}) \\ &\quad + (A_{12} Z_{j-1}^\top + Z_{j-1} A_{21}) A_{21}^\top - A_{21}^\top A_{22} + A_{11} A_{21}^\top \\ A_{11}^\top Z_j - Z_j A_{22}^\top &= X_j \end{aligned} \tag{4.27}$$

starting from e.g. $Z_0 = 0$. This system of linear equations is uniquely solvable if and only if the block matrices A_{11} and A_{22} have disjoint spectra. Once again, we immediately get the following result. Recall, that an invariant subspace V of a linear operator

$A : X \rightarrow X$ is called stable, if the restriction $A|_V$ and corestriction operators $A|_{X/V}$, respectively, have disjoint spectra.

Theorem 4.4 *Algorithm 3 converges locally quadratically to projectors onto stable invariant subspaces of A .*

5 Conclusions

We presented a new differential geometric approach to Newton algorithms on a Graßmann manifold. Both the classical Graßmannian as well as the Lagrange Graßmannian are considered. The proposed Newton algorithms depend on the choice of a pair of local coordinate systems having equal derivatives at the base points. Using coordinate charts defined by the Riemannian normal coordinates and QR -factorizations, respectively, leads to an efficiently implementable algorithm. Using the proposed method, new algorithms for symmetric eigenspace computations and non-symmetric invariant subspace computations are presented that have potential for considerable computational advantages, compared with previously proposed methods.

References

1. P.-A. Absil. *Invariant Subspace Computation: A Geometric Approach*. PhD thesis, University of Liege, Belgium, 2003.
2. P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80:199–220, 2004.
3. R.L. Adler, J.-P. Dedieu, J.Y. Margulies, M. Martens, and M. Shub. Newton’s method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, 2002.
4. A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998.
5. D. Gabay. Minimizing a differentiable function over a differentiable manifold. *J. of Optimization Theory and Applications*, 37(2):177–219, 1982.
6. U. Helmke and J.B. Moore. *Optimization and Dynamical Systems*. CCES. Springer, London, 1994.
7. K. Hüper and J. Trumf. Newton-like methods for numerical optimization on manifolds. In *Thirty-Eighth Asilomar Conference on Signals, Systems & Computers*, pages 136–139, Pacific Grove, California, November 7 – November 10 2005.
8. J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer, Berlin, 2nd edition, 1998.
9. A.W. Knap. *Lie groups beyond an introduction*. Birkhäuser, Boston, 1996.
10. S. Lang. *Fundamentals of Differential Geometry*. Springer, New York, 1999.
11. D. G. Luenberger. *Linear and nonlinear programming*. Addison-Wesley, Reading, 2nd edition, 1984.
12. A. Machado and I. Salavessa. Grassman manifolds as subsets of Euclidean spaces. In L.A. Cordero, editor, *Differential Geometry*, volume 131 of *Research Notes in Mathematics*. Pitman, 1985.
13. R. Mahony. *Optimization algorithms on homogeneous spaces*. PhD thesis, Australian National University, Canberra, March 1994.
14. M. Shub. Some remarks on dynamical systems and numerical analysis. In *Dynamical systems and partial differential equations (Caracas, 1984)*, pages 69–91. Univ. Simon Bolivar, Caracas, 1986.
15. S.T. Smith. Optimization techniques on Riemannian manifolds. In A. Bloch, editor, *Hamiltonian and gradient flows, algorithms and control*, Fields institute communications, pages 113–136. American Math. Soc., Providence, 1994.
16. C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers, Dordrecht, 1994.